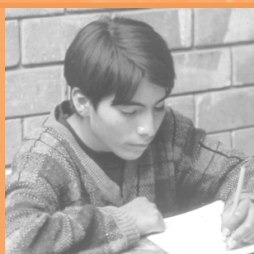




Unidad de Medición
de la Calidad Educativa

UMC



Evaluación Nacional del Rendimiento Estudiantil 2004

Informe descriptivo de resultados

12

Documento de trabajo
UMC

MINISTRO DE EDUCACIÓN

Javier Sota Nadal

SECRETARIO DE PLANIFICACIÓN ESTRATÉGICA

Enrique Prochazka Garavito

SECRETARIO ADJUNTO DE PLANIFICACIÓN ESTRATÉGICA

Walter Twanama Altamirano

JEFE DE LA OFICINA DE PLANIFICACIÓN ESTRATÉGICA

Carlos Pizano Paniagua

JEFA DE LA UNIDAD DE MEDICIÓN DE LA CALIDAD EDUCATIVA

Liliana Miranda Molina

COORDINADORA DEL EQUIPO DE EVALUACIÓN

Tania Pacheco Valenzuela

EQUIPO DE REDACCIÓN DEL INFORME

Andrés Burga León

Liliana Miranda Molina

Tania Pacheco Valenzuela

EDITORIA

Ximena Urbina Keller

© Ministerio de Educación del Perú, 2005.
Calle Van de Velde N° 160, Lima 41 – Perú
Teléfono: 215 5800
www.minedu.gob.pe

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2006-6512

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

Contenido

Introducción	5
1. La medición del rendimiento estudiantil	9
2. La Evaluación Nacional 2004	21
3. Marco conceptual de las pruebas de rendimiento de la EN 2004	48
4. Sistema educativo peruano	56
5. Resultados de la Evaluación Nacional 2004	65
6. Evolución del rendimiento	82
Anexos	89
Anexo 1: Diseño muestral de la EN 2004	91
Anexo 2: Diseño de cuadernillos de las pruebas aplicadas en la EN 2004	105
Anexo 3: Proceso de corrección múltiple de los ítems de producción de respuesta	110
Anexo 4: Establecimiento de puntos de corte	119
Anexo 5: Tablas de resultados generales	122
Conclusiones	149
Bibliografía	153

El proceso de modernización educativa iniciado en nuestro país en la década pasada no solo ha supuesto la redefinición gradual del enfoque que sustenta la práctica pedagógica —del paradigma de la enseñanza al paradigma del aprendizaje—, sino también la implementación de cambios sustantivos en materia de currículo escolar, capacitación docente, materiales educativos e infraestructura educativa. Este proceso, además, ha puesto de manifiesto la necesidad de desarrollar un sistema de evaluación de la calidad que ofrezca información relevante para la toma de decisiones en materia de políticas educativas y que, a su vez, mantenga a la comunidad educativa y a la sociedad en su conjunto informada respecto de los procesos y resultados de nuestro sistema educativo.

En este marco se creó la Unidad de Medición de la Calidad Educativa (UMC) como la instancia técnica del Ministerio de Educación responsable de crear y consolidar un sistema de evaluación de la calidad que ofrezca información relevante respecto de los aprendizajes del sistema educativo peruano, así como de los factores que están asociados a los mismos. Con ello se busca tanto contribuir a una adecuada toma de decisiones en materia de políticas educativas, como satisfacer las demandas de información de la sociedad en su conjunto y de la comunidad educativa en el marco de una política de rendición de cuentas del sector público.

Como parte de su labor, la UMC ha realizado cuatro evaluaciones del rendimiento escolar a escala nacional: CRECER 1996, CRECER 1998, la Evaluación Nacional 2001 y la Evaluación Nacional 2004.¹ Asimismo, desde su creación, la UMC viene participando en estudios internacionales de evaluación educacional como el Primer Estudio Internacional Comparativo de Lenguaje, Matemática y Factores Asociados (1997), realizado por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), órgano técnico de la Oficina Regional de Educación para América Latina y el Caribe (OREALC) de la UNESCO, y el Programa para la Evaluación Internacional de Alumnos (PISA), organizado por la Organización para la Cooperación y el Desarrollo Económico (OCDE).

Cada evaluación ha supuesto un aprendizaje que ha permitido enriquecer y mejorar la formulación de los instrumentos de evaluación del rendimiento, así como realizar análisis cada vez más complejos de los factores que se asocian con los resultados de los estudiantes. Entre estos factores se han considerado procesos escolares, procesos de aula, características de los principales actores del sistema (estudiantes, profesores, directores, padres de familia), entre otras muchas variables y dimensiones que de una u otra forma ofrecen información sobre la calidad del sistema.

1. El diseño, los instrumentos utilizados y las conclusiones y resultados de las distintas evaluaciones se encuentran publicados y disponibles en la página Web de la UMC: www.minedu.gob.pe/umc/

En noviembre del año 2004 se realizó la cuarta Evaluación Nacional del Rendimiento Estudiantil (EN 2004) con la finalidad de proporcionar información a escala de sistema sobre el grado de desempeño que los estudiantes peruanos demuestran respecto a las principales competencias de las áreas de Comunicación y Matemática establecidas en el currículo vigente. Adicionalmente se recogió información sobre el desempeño en algunas competencias del eje curricular de Formación Ciudadana.

Las pruebas de rendimiento aplicadas a los estudiantes han sido diseñadas para identificar qué es lo que los estudiantes saben y hacen respecto de lo que deberían saber y hacer en el grado de estudios que cursan.² En tal sentido, la pruebas de la EN 2004 han permitido recoger información relevante y compleja que apunta a describir los niveles de desempeño sobre lo que esencialmente se quiere desarrollar en los estudiantes: capacidad de analizar, inferir y resolver problemas. Las pruebas también permiten la comparación relativa entre los distintos grupos estudiados de alumnos.

Los grados que se evaluaron fueron segundo y sexto grados de educación primaria y tercero y quinto grados de educación secundaria. Aproximadamente se evaluó a 14 000 estudiantes por grado en 843 instituciones educativas de educación primaria y 636 instituciones educativas de educación secundaria. Al igual que en las evaluaciones anteriores, se aplicaron cuestionarios y fichas de observación a distintos actores de la comunidad educativa: estudiantes, docentes y padres de familia.

El diseño de la muestra fue representativo a nivel nacional, área de residencia (urbano/rural), gestión del centro educativo (estatal/no estatal) y característica del centro educativo en el caso del nivel primario (polidocente completo/ multigrado). En una muestra independiente, se aplicaron pruebas de comprensión de textos y producción escrita en quechua (Cusco-Collao) y aimara como lenguas maternas y en castellano como segunda lengua a estudiantes del sexto grado de primaria de las instituciones educativas del Programa de Educación Bilingüe Intercultural.

El presente informe presenta los resultados que obtuvieron los estudiantes en la EN 2004³ y está organizado en siete capítulos. El primero de ellos presenta un marco general sobre la evaluación del rendimiento estudiantil, la medición y el modelo de medición asumido para el análisis de las pruebas aplicadas, que pertenece a la familia de modelos Rasch.

El segundo capítulo nos describe las características principales de la EN 2004: los grados y áreas evaluadas, el proceso de elaboración de los marcos de evaluación y de aplicación de las pruebas y de los instrumentos de factores asociados, etc. Además contiene una descripción de los niveles de desempeño de todos los grados y áreas evaluados.

El tercer capítulo muestra las dimensiones del modelo general de evaluación asumido. Se especifica que cada una de las preguntas aplicadas en la EN 2004 obedece a una capacidad, un contenido y un contexto. Aquí el lector podrá encontrar la definición que se adoptó para cada uno de estos términos; además se describen los contenidos de estas tres dimensiones tanto para la prueba de Comprensión de textos escritos como para la de Matemática.

En el capítulo cuarto se presenta la situación del sistema educativo a través de algunos indicadores clásicos como cobertura, eficiencia interna, conclusión, etc., todos ellos ana-

2. Es decir, se ha asumido una interpretación de los resultados referida al criterio.

3. Este informe debe ser leído conjuntamente con los Informes pedagógicos de resultados de la EN 2004, así como con el informe sobre el modelo de factores asociados al rendimiento estudiantil.

lizados por diversos ejes de equidad con el fin de contextualizar los resultados que obtienen los estudiantes.

En el capítulo quinto se presentan los resultados que obtuvieron los estudiantes según los niveles de desempeño establecidos. También se encuentra aquí el análisis de las diferencias del rendimiento académico según los estratos considerados en la muestra de estudiantes evaluados.

En el sexto capítulo se describe la evolución del rendimiento académico, mediante la comparación de los resultados promedio en los años 1998 y 2004 —la prueba CRECER 98 fue aplicada en el año 2004 a una submuestra de estudiantes. En este capítulo se encuentra también una explicación de la metodología empleada a fin de hacer ambas aplicaciones comparables.

Finalmente, se presentan las conclusiones a las que se llegaron, luego de analizar la información recogida.



1.1. La evaluación del rendimiento estudiantil

La Unidad de Medición de la Calidad Educativa (UMC) considera que la evaluación es sumamente importante dentro del ámbito de la educación, en la medida en que esta proporciona información valiosa para mejorar el proceso educativo.

La evaluación del sistema educativo es una tarea compleja que compromete muchas dimensiones y que puede ser enfocada desde diversas perspectivas y metodologías. Por ejemplo, se puede evaluar al sistema preguntándose si hay escuelas suficientes para atender a todos los niños en edad escolar, si la infraestructura de sus locales es apropiada y segura, si el presupuesto que maneja es sostenible y suficiente, si la participación de la familia en el proceso educativo es acogida y alentada, si las poblaciones tradicionalmente excluidas forman ahora parte del sistema, si las Unidades de Gestión Local funcionan apropiadamente, etc. Entre estas muchas preguntas, aparece sin duda una que es central: ¿Aprenden los estudiantes lo que el sistema establece que deben aprender?

Se espera que un sistema educativo de calidad logre, entre otras cosas, que los estudiantes que pasan por él se desenvuelvan como sujetos activos, competentes y comprometidos con su desarrollo personal y con el de su entorno. Tal propósito demanda el desarrollo de una serie de capacidades y competencias recogidas en las intenciones educativas y expresadas en el currículo. La evaluación de dichas competencias y capacidades está lejos de ser una tarea sencilla, mucho más si el esfuerzo debe hacerse a escala nacional.⁴

Una forma de aproximarse a gran escala a lo aprendido por los estudiantes es por medio de la evaluación de su desempeño o rendimiento. Para ello se aplican pruebas estandarizadas a muestras representativas de estudiantes cuyo propósito central es responder a la siguiente pregunta: ¿Qué logros de aprendizaje ha adquirido el estudiante tras un periodo de instrucción?⁵ (Prieto y García 1996). Basándonos en la propuesta de Pizarro, Clark y Allen (1987), la medición del rendimiento académico puede ser entendida como la determinación de una cantidad que estima lo que una persona ha aprendido como consecuencia de un proceso de instrucción o formación; es la capacidad del alumno para responder al proceso educativo en función a metas plasmadas en el currículo.

García y Prieto (1996) sostienen que es muy importante que al construir una prueba para evaluar el rendimiento se defina adecuadamente el dominio o conjunto de indicadores a partir de los cuales se infiere el nivel de logro de las personas en aquella materia que se

4. Sobre las características que comporta una evaluación a escala de sistema se puede consultar: «Evaluación de sistema y evaluación de aula», disponible en: http://www.minedu.gob.pe/umc/otros/tipoeval/ev_sistema_ev_aula_preliminar.pdf

5. Thorndike (1989) señala que el método centrado en la evaluación de los logros en algún área del conocimiento corresponde a las llamadas pruebas de aprovechamiento con referencia al criterio.

quiere evaluar. De manera análoga para la evaluación de una variable compleja —como puede ser la competencia lectora, por ejemplo— es importante hacer un análisis del contenido de la variable y derivar del mismo el conjunto de indicadores suficientes para dar cuenta del nivel de desempeño en la variable evaluada. Así, tal como los autores citados establecen, una prueba de rendimiento queda compuesta por indicadores que se conectan con ítemes o tareas significativas asociadas a la competencia por medio de definiciones semánticas. Esto constituye la matriz que servirá como base para la construcción de los ítemes. Dicha matriz suele ser conocida en nuestro medio como las especificaciones de las pruebas.

En nuestro caso, los ítemes —o preguntas— de las pruebas aplicados en la Evaluación Nacional 2004 (EN 2004) fueron seleccionados para ser representativos de las variables en evaluación, es decir, de las competencias o capacidades de las que se quería dar cuenta. Dichas variables fueron a su vez seleccionadas del currículo oficial vigente en razón de su relevancia y de la posibilidad de ser evaluadas mediante pruebas estandarizadas de lápiz y papel.

De ninguna manera planteamos que las pruebas estandarizadas son las únicas que deben utilizarse. Sin embargo, consideramos que son pertinentes para obtener un diagnóstico bastante preciso de los logros mostrados por diversos grupos de estudiantes. Esta evaluación está orientada a generar dicho diagnóstico y no se centra en los logros de algún estudiante en particular o de determinadas instituciones educativas.⁶ Lo que nos interesa es dar información a escala nacional sobre los resultados del proceso educativo.

1.2. Procesos de medición⁷

Stevens (1957) afirma que medir, en un sentido amplio, es asignar numerales a objetos o eventos de acuerdo a reglas (citado en Nunnally y Bernstein 1995). Es decir, medir consiste en establecer reglas para asignar símbolos a objetos de tal manera que:

- representen cantidades o atributos de forma numérica, que indican qué tanto del atributo está presente en el objeto;
- definan si los objetos caben en las mismas categorías o en otras diferentes con respecto a cualidades esenciales.

Podemos distinguir dos tipos de procesos de medición: los directos y los indirectos. En los primeros, se pone en correspondencia directa un instrumento de medida con la propiedad del objeto medido. Por ejemplo, si queremos saber la longitud de una pieza de madera, ponemos en correspondencia uno de sus lados con una regla y haremos afirmaciones como: «Esta pieza mide 27 cm. de largo.» Esto se puede lograr gracias a que la escala de medición está contenida en el instrumento de medida.

Sin embargo, muchas de las características que evaluamos no son directamente observables, son más bien construcciones hipotéticas (rasgos latentes). Diversos autores

6. Cabe precisar que a pesar que las pruebas de la EN 2004 están diseñadas para dar información a escala nacional, el año pasado la UMC le hizo llegar a cada institución educativa un informe con los resultados que han obtenido sus estudiantes evaluados a fin de apoyar procesos informados de reflexión y toma de decisiones de los actores involucrados para mejorar la calidad de los procesos educativos en cada una de estas instituciones.

7. Los apartados que se presentan a continuación en este capítulo constituyen un resumen del artículo de A. Burga (2005).

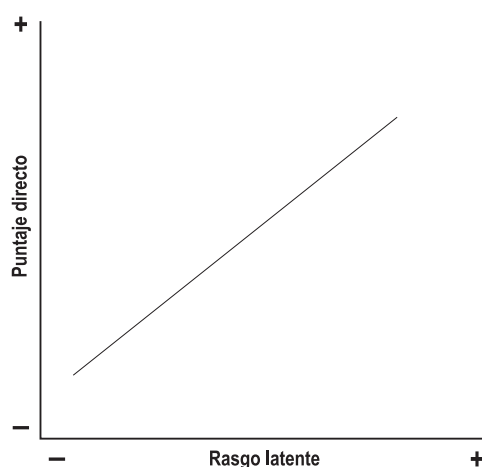
(Muñiz, 1996; Nunnally y Bernstein, 1995; Thorndike, 1989) sostienen que los rasgos latentes son cuantificables, o sea, que tienen propiedades de cantidad o grado, en el sentido de que una persona puede tener más del rasgo que otra, o que una persona tiene más del mismo en un momento dado que en otro. Por ejemplo, un alumno puede estar más motivado que otro por los contenidos de la clase de matemática. Asimismo, mientras que un mismo alumno puede estar muy motivado por esos temas en primaria, puede perder la motivación en secundaria. En este caso la motivación sería un rasgo latente o constructo, puesto que su existencia es inferida a partir del comportamiento del estudiante; por ejemplo, si presta atención a la clase o si manifiesta agrado por la misma se inferirá que está motivado. La motivación, sin embargo, no es observable directamente como lo son, por ejemplo, la estatura o el peso de un objeto. De este modo, la medición de estos rasgos latentes es indirecta, es decir, se realiza a través de expresiones o supuestas manifestaciones del rasgo.

Tal es el caso de las pruebas de rendimiento. Por ejemplo, en nuestro caso las pruebas buscan evaluar el grado o nivel de desarrollo de una competencia determinada en un estudiante. Dicho nivel no es observable ni directamente medible y se infiere a partir del desempeño de un estudiante en una evaluación que le plantea un conjunto de tareas específicas cuya resolución requiere de determinado grado de competencia. Así, el grado de competencia se infiere a partir del desempeño del estudiante en la prueba.

Ahora bien, respecto a cómo han de ser interpretados y analizados los resultados de una prueba existen teorías distintas. Tradicionalmente, en el marco de la Teoría Clásica de los Tests se empleaba como principal referente el puntaje directo. Este se determina, generalmente, como la sumatoria de las puntuaciones obtenidas en cada ítem o pregunta de una prueba. Por ejemplo, si un alumno resuelve correctamente 15 preguntas, y cada pregunta vale dos puntos, su puntaje directo es 30.

De acuerdo con la Teoría Clásica de los Tests (TCT), la relación entre los puntajes directos de las pruebas y la cantidad del rasgo latente es monotónica⁸ lineal (Muñiz 1996). Lo que quiere decir que iguales cantidades de aumento en el puntaje directo corresponden a aumentos de la misma magnitud en el rasgo latente. El gráfico 1.1. ilustra este tipo de relación.

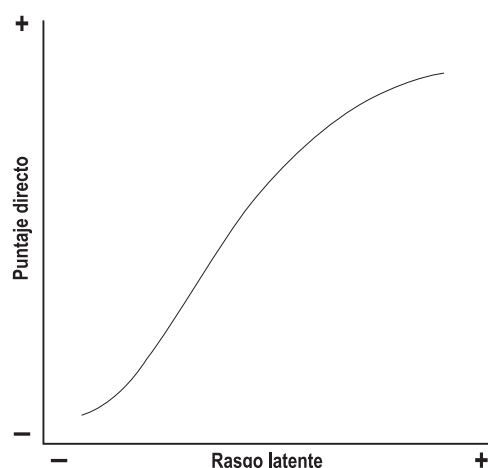
Gráfico 1.1: Relación no lineal entre el rasgo latente y el puntaje directo en la Teoría Clásica de los Test



8. Se plantea que una relación es monotónica cuando un aumento de los valores en una de las variables se encuentra asociado al aumento de los valores de la otra variable.

Sin embargo modelos más actuales, como la Teoría de Respuesta al Ítem (TRI), sostienen que la relación entre los puntajes directos de las pruebas y la cantidad del rasgo latente podría ser caracterizada más bien como monotonía no lineal (Hambleton, Swaminathan y Rogers 1999). Es decir, la cantidad del rasgo latente que se necesita, por ejemplo, para pasar de un puntaje directo de 2 puntos a otro de 4 no es la misma que se necesita para pasar de 15 a 17 puntos.⁹ Como apreciamos en el siguiente gráfico, aumentos de la misma magnitud en el puntaje directo no corresponden a la misma magnitud de aumento en el rasgo latente:

Gráfico 1.2: Relación no lineal entre el rasgo latente y el puntaje directo (TRI)



Así, la TRI sostiene que la relación entre el desempeño del examinado y las habilidades que la sustentan puede describirse por una función monotonía creciente llamada «curva característica del test». Esto último implica que mientras mayor sea la habilidad de una persona, es mayor la cantidad de ítemes respondidos de forma correcta. Si bien la relación antes señalada entre la cantidad del rasgo latente y el número de ítemes correctos es sostenida por la TRI, las relaciones entre habilidad y desempeño se modelan a nivel de ítemes, mediante las curvas características del ítem. Dichas curvas nos muestran la relación existente entre la probabilidad de responder correctamente a un ítem y la cantidad de rasgo latente que posee una persona, como lo veremos en detalle en las siguientes secciones.

1.3. Teoría de Respuesta al Ítem

El nombre proviene del hecho de que estos modelos se centran fundamentalmente en las propiedades de los ítemes o preguntas. Concretamente, tratan de modelar matemáticamente¹⁰ qué ocurre cuando una persona con una habilidad determinada se enfrenta a un ítem específico.

9. Desde esta perspectiva, entonces, las puntuaciones directas no constituyen una verdadera escala de intervalo. Sin embargo, existe un conjunto de modelos matemáticos que sirven para linealizar dichas relaciones, construyendo medidas a partir de los puntajes directos, de tal manera que tengan las propiedades de una escala de intervalo. Estos modelos matemáticos en general se conocen como Teoría de Respuesta al Ítem.

10. Como se verá más adelante, estos modelos matemáticos son probabilísticos.

Una gran ventaja de estos modelos es que dan la posibilidad de tener mediciones invariantes respecto de los instrumentos utilizados y las personas evaluadas. Es decir, los parámetros de los ítemes o preguntas son invariantes aunque se estimen en diferentes muestras de la población. Este es un problema central en Teoría Clásica de los Tests. En ella, por ejemplo, si un alumno responde a tres ítemes difíciles, obtendrá un puntaje directo igual al de otro alumno que haya respondido a tres ítemes fáciles. Un alumno que responde a muchas preguntas fáciles tendrá una puntuación directa alta en comparación con otro que responde a unas pocas preguntas difíciles. Erróneamente, se podría pensar que el alumno con un puntaje directo superior posee más habilidad. Sin embargo, no se estarían considerando los efectos de la dificultad de los ítemes (y de la prueba en general) para establecer el puntaje de un alumno en particular. Un alumno puede tener un puntaje directo más bajo en un test difícil que el obtenido por otro alumno en un test fácil, y, a pesar de ello, tener una habilidad más alta.

Además, los parámetros obtenidos para describir los ítemes en Teoría Clásica de los Tests dependen de la muestra de personas en las cuales fueron calculados. Por ejemplo, el típico índice de dificultad para ítemes dicotómicos se obtiene al dividir el número de aciertos al ítem entre la cantidad de personas a las que se les presentó el mismo. Si ese ítem es respondido por personas con alta habilidad, la proporción de aciertos (índice p) será alta, por lo tanto el ítem parecerá fácil. Por otro lado, si ese mismo ítem es respondido por una muestra de personas con baja habilidad, la proporción de aciertos será baja y el ítem parecerá difícil. El índice p únicamente nos dice cuán difícil fue ese ítem para el grupo de personas en el que fue calculado.

Frente a estos problemas encontrados en la Teoría Clásica de los Tests, surgen los modelos de Teoría de Respuesta al Ítem que permiten (Hambleton, Swaminathan y Rogers 1991):

- obtener mediciones que no varíen en función del instrumento utilizado, es decir, que presenten invarianza respecto de los ítemes empleados; y
- disponer de ítemes cuyos parámetros no dependan de los objetos medidos, es decir, que presenten invarianza respecto de las personas evaluadas.

A parte de la invarianza, hay dos requerimientos muy importantes en los ítemes que constituyen un test o prueba con el fin de aplicar los modelos de Teoría de Respuesta al Ítem: la unidimensionalidad y la independencia local.

La unidimensionalidad implica que un solo rasgo latente o habilidad se encuentra en la base de un conjunto de ítemes. En otras palabras, un instrumento será unidimensional si las respuestas dadas a él son producidas en base a un único atributo. Wright y Linacre (1998) señalan que, en la práctica, ningún instrumento puede ser perfectamente unidimensional. Lo que buscamos es tener instrumentos que, en esencia, muestren unidimensionalidad. Lo importante es que un instrumento de medida mida fundamentalmente un solo atributo.

En el marco de la Teoría Clásica de los Tests, se asume que, a mayor puntaje directo, más de ese rasgo o atributo posee la persona evaluada. ¿De dónde proviene ese puntaje directo o puntaje global? Proviene de la suma de los puntajes obtenidos en cada uno de los ítemes. Obtener los puntajes totales sumando las calificaciones de cada ítem supone que se está midiendo con ellos un solo atributo, de lo contrario no habría ningún fundamento que soporte las operaciones aritméticas realizadas con los ítemes. De la misma manera, si se pretende medir la cantidad de una variable, esta no debe estar *contaminada*

por las cantidades que posee la persona evaluada en otras variables (Stout 1987; citado en Cuesta 1996).

El otro requerimiento importante que señalamos es la independencia local. Esta implica que al mantener constante el rasgo latente causante de las respuestas a los ítems de un test, las respuestas de los examinados a cualquier par de ítems son estadísticamente independientes (Hambleton, Swaminathan y Rogers 1991). La probabilidad de responder correctamente a un ítem no está relacionada con la probabilidad de responder correctamente a otro ítem siempre y cuando la habilidad del examinado sea controlada. Es decir, asumiendo tal control, la probabilidad de responder correctamente tanto al ítem uno como al dos es igual al producto de la probabilidad de responder correctamente a cada uno de los ítems. Esto implica que no hay ningún tipo de dependencia en los ítems aparte de la atribuible al rasgo latente evaluado. Por ello, Embretson y Reise (2000) señalan que se logra la independencia local cuando las relaciones observadas entre los ítems son explicadas casi en su totalidad por un modelo de Teoría de Respuesta al Ítem.

Cuando el conjunto de ítems utilizados se comporta unidimensionalmente, se obtiene paralelamente la independencia local (Hambleton, Swaminathan y Rogers 1991). Como lo señalan Embretson y Reise (1991), la unidimensionalidad implica independencia local, pero la independencia local no necesariamente implica unidimensionalidad. En otras palabras, un conjunto de ítems multidimensionales puede también mostrar independencia local.

1.4. Modelos Rasch

En la UMC se ha utilizado el modelo Rasch como insumo para el análisis de las pruebas aplicadas en la EN 2004. Este modelo, que es considerado por algunos como un tipo de modelo de TRI,¹¹ se centra en el análisis de cada ítem, concretamente en la interacción entre una persona y un ítem. Establece la probabilidad de respuesta de una persona ante un ítem en términos de la diferencia entre la medida de rasgo o habilidad latente de la persona (B) y la medida del ítem utilizado en términos de su dificultad (D)¹².

George Rasch, matemático danés, se dio cuenta de que los resultados de la interacción entre personas e ítems no pueden estar totalmente predeterminados, sino que estos siempre comportan un elemento de impredecibilidad (Wright y Linacre 1989). Esto conlleva al requerimiento de que, en términos probabilísticos, mientras mayor habilidad, mayor 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.

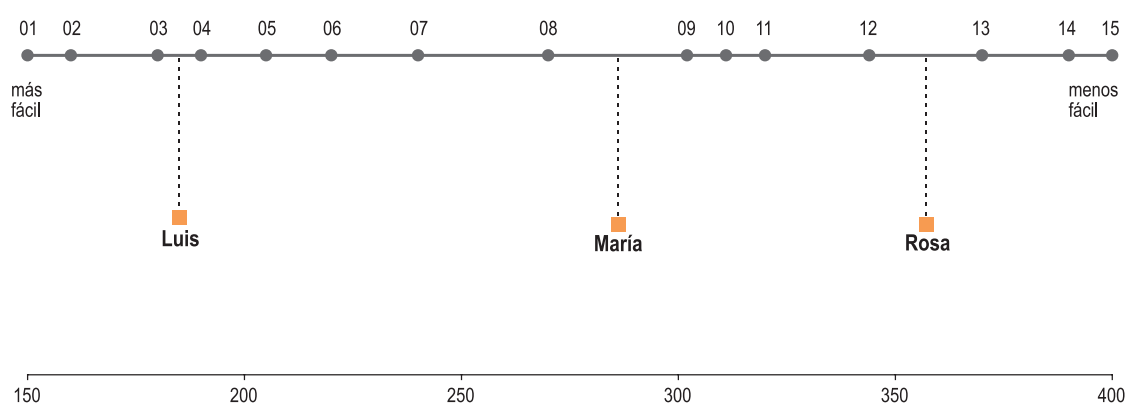
11. Al respecto, existe una controversia entre los seguidores del análisis Rasch y los de los modelos de TRI. En general, el análisis Rasch plantea que sus modelos son una definición de medición (Wright 1989). Es decir, señalan que lo importante es ver hasta qué punto los datos se ajustan a dicha definición. Si se da este ajuste, se habrá construido una buena medida. En cambio, la TRI se orienta a encontrar qué modelo matemático se ajusta mejor a los datos. Más información sobre esta controversia puede encontrarse en Shaw (1991).

12. En algunos textos se utiliza la letra theta (θ) para referirse a la habilidad de las personas y la letra b para referirse a la dificultad de los ítems.

En los modelos Rasch la habilidad de las personas y las dificultades de los ítems se ubican en la misma métrica. Ello permite combinar la habilidad de la persona y la dificultad del ítem para predecir el desempeño de una persona en un ítem cualquiera e identificar respuestas inesperadas. La idea central del análisis Rasch es 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):

Gráfico 1.3: Representación lineal de la dificultad de los ítems y habilidad de las personas



Si sabemos que Luis tiene una habilidad mayor que la dificultad del ítem 03, pero menor que la 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 afirmado que es probable que haya acertado todos los más fáciles y fallado los más difíciles. No estamos afirmando que siempre 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 ítems 07 y 08, fallado el 09, acertado el 10 y fallado todos los demás.

1.4.1. MODELO PARA ÍTEMES DICOTÓMICOS

Un ítem dicotómico tiene una sola respuesta correcta, por lo tanto, al responderlo, se puede acertar y recibir un punto ($X=1$) o fallar y no recibir ningún puntaje ($X=0$).

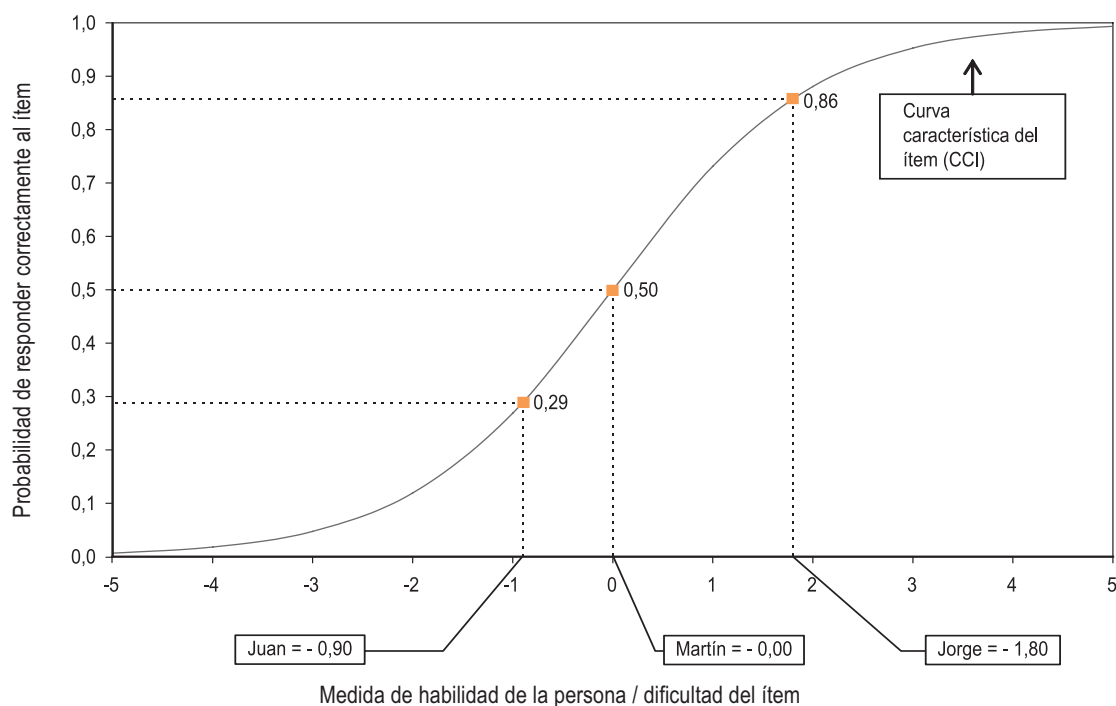
La relación entre la habilidad y dificultad puede graficarse por medio de las curvas características del ítem (CCI) que nos dan información concreta sobre la probabilidad de

respuesta de una persona ante un ítem. En el caso de tener ítems dicotómicos, al trazar dichas curvas se presentan las siguientes relaciones:

1. $B > D$ (la habilidad (B) es mayor que la dificultad (D) del ítem)
2. $B < D$ (la habilidad (B) es menor que la dificultad (D) del ítem)
3. $B = D$ (la habilidad (B) es igual a la dificultad (D) del ítem)

El primer caso nos muestra que, si la habilidad de la persona es mayor a la dificultad del ítem, la probabilidad de responder correctamente a dicho ítem y recibir un punto es mayor que 0,50 (50%).¹³ 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 y recibir un punto es menor que 0,50 (50%).¹⁴ 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 0,50 (50%).¹⁵ Esta relación entre la habilidad de la persona y la dificultad del ítem puede graficarse de la siguiente manera:

Gráfico 1.4: Curva característica de un ítem ficticio y probabilidad de cuatro alumnos con habilidades diferentes de responderlo de forma correcta

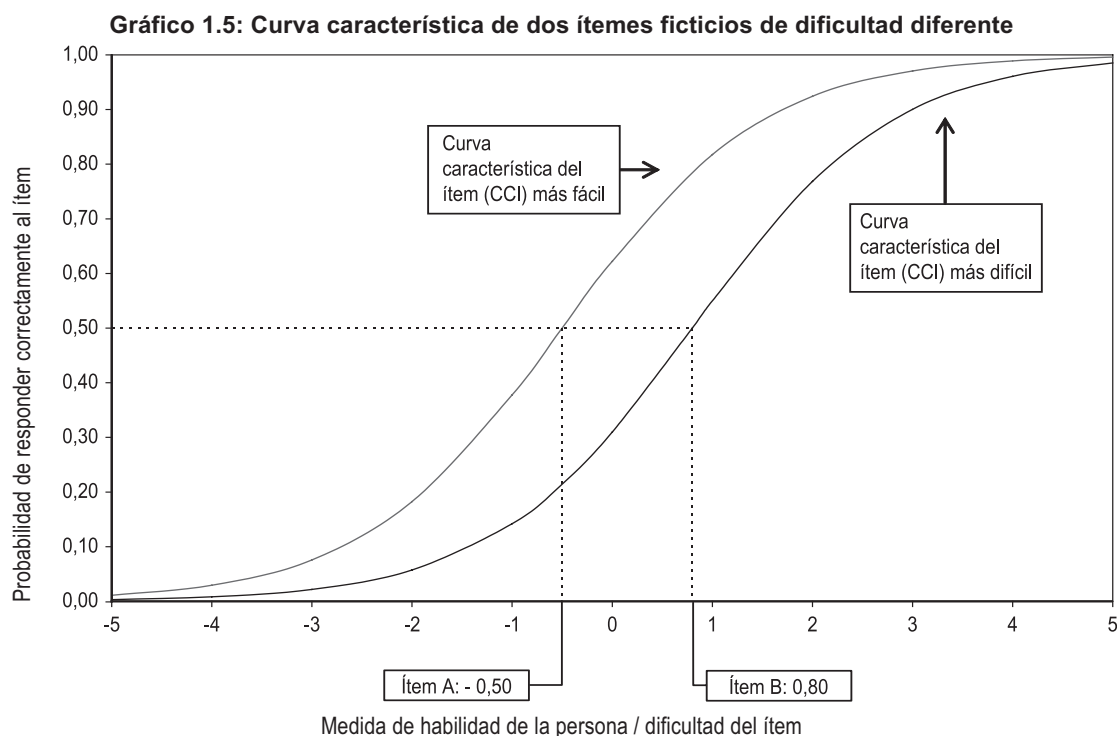


En el gráfico anterior se ve que, al enfrentarse a este ítem en particular, Juan, cuya habilidad respecto a este ítem es de -0,90, tiene una probabilidad de 0,29 de acertarlo, es decir, lo más probable es que lo falle y obtenga 0 puntos. En cambio, Jorge, cuya habilidad respecto a este ítem se ha estimado en 1,80, tiene una probabilidad de 0,89 de responderlo correctamente, 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.

13. Si $B > D \rightarrow P(x=1 / B,D) \in]0,50 ; 1,00]$

14. Si $B < D \rightarrow P(x=1 / B,D) \in [0,00 ; 0,50[$

15. Si $B = D \rightarrow P(x=1 / B,D) = 0,50$

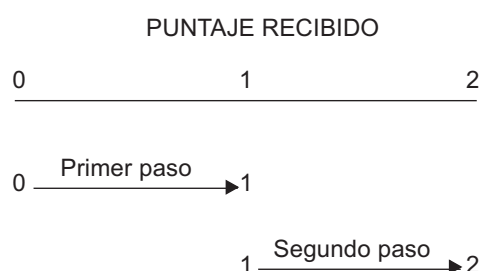


Al comparar dos o más curvas características de ítems, se puede decir que, mientras más a la derecha se encuentra una curva, más difícil es el ítem. Además, se expresa numéricamente la dificultad de un ítem como aquel valor de la habilidad que posee una probabilidad de 50% de acertar a dicho ítem. Así, en el Gráfico 1.5, el ítem más fácil tiene una dificultad de -0,50 y el más difícil de 0,80.

1.4.2. MODELO DE CRÉDITOS PARCIALES

Masters introduce en 1982 el Modelo de Créditos Parciales para trabajar con ítems politómicos de categorías ordenadas (Verhelst y Verstralen 1997). El modelo especifica que cada ítem tiene su propia estructura de calificación. Se deriva de los test de opción múltiple en los que hay respuestas incorrectas, pero que indican algún conocimiento, a las que se les otorga un crédito parcial (Wright 1999). Fox (1999) señala que el Modelo de Créditos Parciales es una generalización del modelo Rasch que se puede aplicar en situaciones en las cuales los ítems pueden variar en el número de alternativas correctas y cantidad de opciones de respuesta en un mismo test.

Al plantear este modelo, se supone que, en el proceso de resolución de un ítem, una persona responde de manera secuencial a un conjunto de subproblemas en el ítem. Los créditos parciales dados equivalen al número de pasos que deben completarse antes de obtener la respuesta final al problema:



Veamos como ejemplo el siguiente ítem del área de Matemática:

Un tren tiene 27 vagones pequeños con 24 pasajeros cada uno. Además hay un vagón grande con 89 pasajeros. ¿Cuántos pasajeros hay en total en el tren?

Escribe aquí tu procedimiento

Respuesta:

0 1 2

Puntaje completo

El estudiante llega a la respuesta 737 (independientemente de la unidad empleada), mostrando un estrategia de solución coherente.

0 1 2

Puntaje parcial

El estudiante comprende la situación planteada y muestra una estrategia conducente la respuesta correcta, es decir, identifica que hay que hallar cuántos pasajeros hay en total en los 27 vagones pequeños (648) y añadirle la cantidad de pasajeros del vagón grande (89), pero comete algún error operativo, llegando a otra respuesta.

0 1 2

Puntaje nulo

El estudiante deja la pregunta en blanco, presenta ideas sin relacionar o llega a otras respuestas que no muestran una estrategia coherente o conducente a la respuesta correcta.

En este modelo se plantea que los puntajes o créditos superiores deben tener una mayor probabilidad de ser alcanzados en ítemes con baja dificultad que en ítemes con alta dificultad. Además, las personas evaluadas con mayores niveles del rasgo latente tienen mayor probabilidad de recibir un crédito total que las personas con menores niveles del rasgo. Ambas propiedades deben mantenerse para todas las personas, todos los ítemes y todas las categorías de respuesta. Asimismo, cuando estas propiedades se cumplen, los puntajes directos obtenidos pueden ser transformados para lograr una escala de medición de intervalo.

Se quiere modelar diferentes niveles de conocimiento frente a un mismo ítem. Es decir, se modela la probabilidad de tener 1 en lugar de 0 puntos en el ítem, y luego la probabilidad de tener 2 en lugar de 1 punto en el ítem. En este modelo puede haber muy pocas o ninguna observación en alguna de las categorías de respuesta de algunos ítemes. Por lo

tanto, las estimaciones de la dificultad de esos pasos no será muy segura (Linacre 2000). Al trabajar con estos modelos es importante asegurar un buen número de respuestas dentro de cada uno de los créditos probables.

1.4.3. AJUSTE AL MODELO Y CONFIABILIDAD

Ningún conjunto de datos se ajusta perfectamente a un modelo matemático, incluyendo la curva normal. Ninguna variable se distribuye exactamente según esta distribución. Lo importante es hasta qué punto es significativo dicho desajuste de los datos con respecto al modelo matemático (Schulz 1990).

Una ventaja importante de los modelos Rasch es que proporcionan medidas de ajuste de los ítems y las personas. Por ejemplo, una persona con desajuste implicaría un patrón inesperado de respuesta, que puede tener diversas explicaciones (responde al azar, falla los ítems fáciles y acierta los difíciles, etc.). Si un ítem muestra desajuste con el modelo, podría explicarse por su falta de discriminación, o porque este ítem está midiendo algo muy distinto al resto de ítems (y por lo tanto no cumpliría con el principio de unidimensionalidad).

Hambleton, Swaminathan y Rogers (1991) sostienen que dos fuentes más del descontento con la Teoría Clásica de los Test descansan en la definición de la confiabilidad y lo que se puede pensar como su inverso conceptual: el error estándar de la medida.

Dentro del modelo de la Teoría Clásica de los Tests, Muñiz (1996) señala que las mediciones 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. Así, el objetivo principal de la confiabilidad es tratar de estimar el error existente en las medidas mediante un indicador denominado coeficiente de confiabilidad.

La confiabilidad, en este marco, se define también como la correlación entre los puntajes del test en formas paralelas de una prueba.¹⁶ Dos pruebas son paralelas cuando miden el mismo rasgo latente con la misma cantidad de ítems, tienen la misma media aritmética y la misma varianza. Si dos test miden lo mismo, cualquier diferencia de puntuaciones en ambos test debe ser producto del error de medición. En la práctica, satisfacer los requerimientos de la definición de test paralelos es difícil, si no imposible. Por otro lado, la confiabilidad es reportada como si fuese una característica invariante, cuando no lo es. Depende no solo del test, sino de la distribución de habilidad de la muestra estudiada y del número de ítems utilizado para evaluarlos (Muñiz 1996).

Mientras más confiable sea un test, menor será el error estándar de medición que este posee. El problema con la medida del error estándar es que se supone que todos los examinados son medidos con la misma precisión, independientemente del nivel de rasgo latente que poseen (Hambleton, Swaminathan y Rogers 1991). Linacre y Wrigth (1989) sostienen que, al ajustar los datos al modelo Rasch para utilizarlos en el establecimiento de medidas, el objetivo es construir un sistema invariante de medidas de intervalo, estimar su precisión (error estándar) y evaluar hasta qué punto estas medidas y sus errores son confirmados por los datos (medidas de ajuste). Una ventaja de los modelos Rasch es

16. Al correlacionar dos pruebas paralelas, teóricamente se debería obtener un coeficiente igual a 1,00, pues al ser paralelas es como si correlacionáramos una variable consigo misma. Sin embargo, la presencia del error de medición hace que las correlaciones con formas paralelas sean diferentes a 1,00.

que permiten calcular un error estándar para cada una de las medidas. En general, se estiman con mayor precisión las medidas cercanas al promedio, mientras que las medidas de los extremos superior e inferior del rasgo latente son estimadas con menor precisión.

Luego de haber visto en estas páginas qué implica medir y cuáles son los dos modelos teóricos que plantean una relación entre la habilidad de la persona evaluada y las respuestas observadas a un conjunto de ítems, nos hemos centrado en la exposición de los principales aspectos de la Teoría de Respuesta al Ítem, de naturaleza probabilística. Esto se ha hecho porque el modelo de Rasch empleado en el análisis de las pruebas aplicadas en la Evaluación Nacional 2004 es un modelo probabilístico.

En el siguiente capítulo se describirán las principales características de la Evaluación Nacional 2004. Se detallarán las áreas y grados evaluados, los estratos representativos de la muestra, los tipos de instrumentos aplicados, los procesos seguidos para la definición de los niveles de desempeño y demás aspectos relevantes del diseño de la evaluación.



2.1. ¿Qué es la Evaluación Nacional 2004?

La Evaluación Nacional 2004 (EN 2004) es la cuarta evaluación nacional del rendimiento estudiantil. Esta evaluación recoge información acerca del sistema educativo peruano en su conjunto mediante la aplicación de diversos instrumentos, como pruebas de rendimiento y cuestionarios, a los diferentes actores del proceso educativo.

Evaluaciones de este tipo se llevan a cabo periódicamente con el fin de proporcionar al sistema educativo, a los investigadores y a la sociedad en general información válida y oportuna sobre el rendimiento académico de los estudiantes y sobre los factores o condiciones escolares y extraescolares que se asocian a este rendimiento. De esta manera, la evaluación identifica los aspectos que deben ser considerados para mejorar los niveles de aprendizaje de los estudiantes peruanos.

Las pruebas de rendimiento aplicadas a los estudiantes en la EN 2004 han sido diseñadas bajo un modelo de evaluación basado en criterios. Este modelo permite identificar qué es lo que los estudiantes saben —y hacen—, respecto de lo que deberían saber y hacer de acuerdo con el grado de estudios que cursan. También permite la comparación relativa entre los distintos grupos. Por ejemplo, se puede comparar el rendimiento de los alumnos de área rurales, frente al de los alumnos de áreas urbanas.

2.2. ¿Qué áreas y grados se evaluaron?

En la EN 2004 han sido evaluadas las áreas curriculares de Comunicación y Matemática,¹⁷ pues proporcionan las herramientas necesarias para el logro de aprendizajes en otras áreas, por lo que se consideran un buen indicador del rendimiento académico global. Adicionalmente, se recogió información para un estudio que tiene el objetivo de aproximarse al eje curricular de Formación Ciudadana.¹⁸

17. Se usará los nombres de Comunicación y Matemática para referirse también a las áreas de Comunicación Integral y Lógico Matemática del nivel primario.

18. Los resultados de la evaluación de este eje se presentan en un informe aparte.

Cuadro 2.1: Grados evaluados según área o eje en la EN 2004

		Matemática	Comunicación	Formación Ciudadana
Primaria	Segundo	✓	✓	—
	Sexto	✓	✓	✓
Secundaria	Tercero	✓	✓	—
	Quinto	✓	✓	✓

Como se aprecia en el cuadro 2.1, se eligieron los grados de sexto de primaria y quinto de secundaria, pues es sumamente importante recoger información sobre los logros alcanzados por los estudiantes al término de cada nivel educativo. Por otra parte se debe considerar que los estudiantes de quinto grado de secundaria no solo están finalizando un nivel, sino que están terminando la escolaridad, por lo que esta evaluación constituye un buen diagnóstico de las capacidades desarrolladas por los estudiantes tras su paso por la Educación Básica Regular. De otro lado, evaluar sexto grado de primaria y quinto de secundaria —poblaciones evaluadas en los años 1998 y 2001— permite proporcionar datos de tendencia en tres puntos de tiempo a lo largo de un periodo de siete años.

Adicionalmente, se eligió evaluar segundo de primaria y tercero de secundaria por ser estos los grados terminales del primer ciclo de sus respectivos niveles.¹⁹

2.3. ¿Quiénes participaron en la EN 2004?

Participaron estudiantes de 843 instituciones educativas (IE) de educación primaria y 636 de educación secundaria de todas las regiones del Perú. Aproximadamente, fueron evaluados 14 500 estudiantes en cada grado.

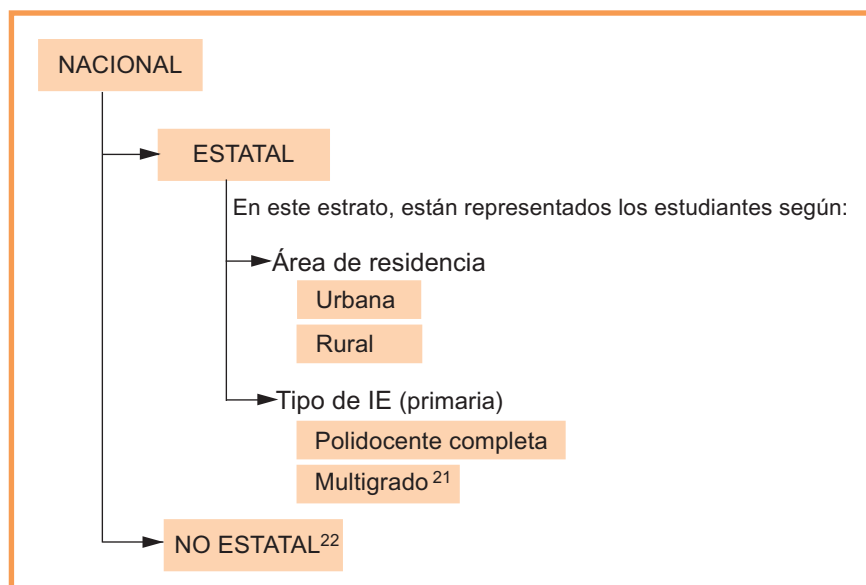
2.4. ¿Cómo se eligieron a las IE participantes?

Se seleccionaron las IE de manera aleatoria (al azar), cuidando que la muestra seleccionada sea técnicamente adecuada para realizar inferencias respecto del rendimiento de toda la población estudiantil del Perú y de los estratos establecidos para los grados evaluados. Una vez definidas las IE participantes, fueron seleccionados también de manera aleatoria el turno, la(s) sección(es) y un total de 30 estudiantes como máximo en cada grado.²⁰

19. Este criterio tuvo como base los currículos vigentes en el momento en que se diseñó la evaluación. El Plan de Estudios de la Educación Básica Regular actual establece que segundo de primaria es el último grado del ciclo III y que tercero de secundaria es el primer grado del ciclo VII, ciclo final de la educación secundaria.

20. Para más detalle véase Anexo 1: Diseño muestral de la EN 2004.

Gráfico 2.1: Estratos representativos de la muestra de la EN 2004



Los recuadros sombreados en el gráfico anterior muestran los estratos representados en la EN 2004. La información presentada en este reporte tiene índices de precisión y confiabilidad aceptables en estos estratos.²³

2.5. ¿Qué instrumentos se aplicaron?

Se aplicaron dos tipos de instrumentos: pruebas de rendimiento y cuestionarios de factores asociados.

- *Las pruebas de rendimiento*

Evalúan el nivel de dominio de las capacidades y los contenidos curriculares. Constan de un conjunto de preguntas para que el estudiante marque o elabora su respuesta. Si bien las pruebas de rendimiento elaboradas para la EN 2004 son de papel y lápiz, presentan una serie de situaciones significativas para los estudiantes con la intención de evaluar el grado de incorporación de los aprendizajes como una herramienta útil para enfrentar diversas situaciones dentro y fuera de la escuela.

- *Instrumentos de factores asociados al rendimiento*

Su objetivo es recoger información que permita contextualizar e identificar aquellos factores que favorecen o perjudican los resultados obtenidos por los estudiantes en las pruebas de rendimiento. Se aplicaron cuestionarios, fichas de observación y otros instrumentos de recolección de datos dirigidos a los distintos actores

21. En este estrato están incluidas las IE Unidocentes.

22. Entre otras IE en este estrato se encuentran las IE cooperativas, parroquiales, particulares, comunales, etcétera.

23. Al respecto consultar el Anexo 5: Cuadros de resultados por estratos y departamentos.

de cada escuela seleccionada: los estudiantes evaluados, sus padres o apoderados, sus profesores y su director.²⁴

2.6. ¿Cómo son las pruebas de rendimiento de la EN 2004?

Las pruebas de rendimiento presentan preguntas de diversos formatos. Entre estos se destacan las preguntas de producción de respuesta llamadas así porque piden a los estudiantes que elaboren o produzcan su propia respuesta. Por ejemplo, se les pidió a los estudiantes que redacten y justifiquen su respuesta o que muestren el procedimiento completo para llegar a esta.²⁵

Considerando el área, las pruebas de rendimiento estaban constituidas por preguntas que han sido elaboradas según los siguientes formatos:

- *En el área de Matemática:* La prueba está constituida por un conjunto de ítemes para marcar la opción correcta, aparear, escribir una respuesta corta (una palabra, un número, etc.), desarrollar el procedimiento necesario para resolver un problema, justificar una afirmación, etc.
- *En el área de Comunicación:* Para evaluar esta área se han elaborado dos pruebas:
 - *Comprensión de textos:* está constituida por diversos textos seguidos de un conjunto de preguntas de diferentes formatos: para marcar la opción correcta, escribir una respuesta corta o desarrollar una respuesta extensa.
 - *Producción de textos:* está constituida por estímulos que invitan a los estudiantes a escribir un texto y, a continuación, un espacio para que el estudiante lo produzca.

El modelo de evaluación empleado en la EN 2004, llamado evaluación basada en criterios, requiere de una gran cantidad de ítemes para poder recoger información sobre lo que saben y hacen los estudiantes respecto de lo que deberían saber y hacer de acuerdo con la estructura curricular. Estos ítemes responden a una propuesta de evaluación en la que las especificaciones de las pruebas determinan el número de ítemes necesarios para evaluar las capacidades más importantes del área. Este número excede lo que podría realizar un estudiante, aún en varias sesiones de evaluación, pues se trata de más de un centenar de preguntas para cada una de las pruebas. Debido a este hecho, se ha recurrido al diseño de bloques balanceados incompletos: los ítemes seleccionados para evaluar un área se dividen en bloques y un estudiante responde solo a algunos de esos bloques, pero no a todos. Luego es posible estimar el desempeño de los grupos de estudiantes en toda la prueba a partir de sus respuestas a un subgrupo de preguntas utilizando la metodología del análisis Rasch.²⁶ Estos diseños de bloques o matrices incompletas se suelen usar en diversos proyectos internacionales de evaluación del rendimiento (por ejemplo, PISA, TIMSS, PIRLS y LLECE) con buenos resultados.

24. Los instrumentos de factores asociados aplicados en la EN 2004 pueden consultarse en: http://www.minedu.gob.pe/umc/eval_nacional2004cf.php.

25. Aproximadamente el 40% de las preguntas de las pruebas aplicadas en la EN 2004 presenta este formato.

26. Para más detalle, véase Anexo 2: Diseño de cuadernillos de las pruebas aplicadas en la EN 2004.

2.7. ¿Cómo se elaboraron las pruebas de rendimiento?

Para la elaboración de las pruebas de rendimiento se ha tomado como base, tanto para seleccionar los contenidos como para determinar las capacidades, los currículos oficiales vigentes de cada uno de los niveles. A continuación se describe el proceso de elaboración de las pruebas:

1. Análisis curricular. Se analizaron no solo los documentos oficiales como la Estructura Curricular Básica (ECB) de Educación Primaria y el Diseño Curricular Básico (DCB) de Educación Secundaria,²⁷ sino también los textos educativos del área de mayor circulación en el medio.
2. Elaboración del Marco de Trabajo. El marco incluye el enfoque del área, las especificaciones de las pruebas —es decir, la selección y adecuación de las capacidades y contenidos a evaluar—, la determinación del número apropiado de preguntas para evaluar cada contenido y los formatos adecuados para estas.
3. Elaboración de las preguntas según las especificaciones de la prueba.
4. Aplicaciones piloto de las preguntas de la prueba.²⁸
5. Análisis estadístico y pedagógico de los resultados de las aplicaciones piloto.
6. Elaboración de las pruebas definitivas.
7. Establecimiento de los puntos de corte de las pruebas.

Todos estos procesos han comprometido la participación de diversos especialistas. Algunos de dichos procesos han requerido de talleres de consulta a pedagogos de las áreas curriculares en evaluación, a docentes con experiencia en la enseñanza a estudiantes de los grados que evalúan las pruebas y a otros profesionales que pudieran aportar desde sus perspectivas a la generación de pruebas apropiadas, no solo en términos estadísticos sino también en términos conceptuales. La construcción de las pruebas de la EN 2004 ha sido, entonces, un ejercicio colectivo desarrollado en múltiples fases interdependientes y coordinado por un equipo multidisciplinario.

Ha sido especialmente importante observar el comportamiento de las preguntas en campo. Las aplicaciones piloto permitieron el acercamiento a los propios estudiantes y la puesta a prueba de los instrumentos. Esto ha permitido, entre otras cosas, evaluar las mejores maneras de formular las preguntas, detectar y relativizar los posibles sesgos culturales o sociales y establecer los tiempos requeridos por los estudiantes para responder a los instrumentos.

2.8. ¿Cuándo y cómo se aplicaron las pruebas?

La aplicación de los instrumentos se llevó a cabo de manera simultánea en todas las IE de la muestra durante la segunda semana del mes de noviembre de 2004. Para ello se tuvo que establecer una red de aplicación tanto para la distribución de los instrumentos como para la selección y capacitación de profesionales de distintas áreas para la admi-

27. Vigentes al momento de diseñar la evaluación.

28. Según las necesidades técnicas de algunas de las pruebas, se llevaron a cabo varias aplicaciones piloto a diversas muestras.

nistración, coordinación, supervisión, control de calidad y aplicación de los instrumentos. Dicha red incluía a dos docentes por cada una de las IE de la muestra.²⁹

Debe precisarse que la decisión de fijar la aplicación de la prueba en noviembre obedece principalmente a la necesidad de garantizar que, para entonces, los estudiantes evaluados hubieran tenido la oportunidad de ser expuestos a la mayoría de las actividades programadas para el año lectivo.

En cada uno de los grados, los estudiantes resolvieron una sola prueba de rendimiento por día. El tiempo máximo con el que contaba cada estudiante para la resolución de cada prueba era de sesenta minutos, con la posibilidad de ser extendido por diez minutos más en caso de que existiera en el grupo evaluado algún alumno que manifestara no haber concluido su prueba.³⁰ La distribución horaria buscaba que el rendimiento de los estudiantes no se viera afectado negativamente por el cansancio.

Como se mencionó, la extensión de las pruebas en cada una de las áreas, es decir, tanto la determinación del número de preguntas, como el tiempo otorgado para la resolución de estas, fue determinada mediante el análisis y revisión de la información recogida en las aplicaciones piloto realizadas en las fases previas.

Para rendir las pruebas en condiciones estandarizadas, cada estudiante recibió un cuadernillo de prueba y una cartuchera con los útiles necesarios para desarrollarla (lápiz, tajador, borrador y regla). Los estudiantes respondieron a las preguntas en el mismo cuadernillo de la prueba. No se usaron hojas de respuesta separadas ni fichas ópticas para evitar que esta condición introdujera dificultad innecesaria a la tarea de responder.

2.9. ¿Cómo se codificaron las respuestas de los estudiantes?

El proceso de codificación de los instrumentos se llevó a cabo por un grupo de docentes especialmente capacitados y seleccionados.³¹ En efecto, como se ha señalado, las pruebas constan no solo de preguntas de opción múltiple, sino de otros formatos en los que el estudiante debe elaborar su propia respuesta. Esto último ha permitido recoger mayor cantidad de información, en especial la relacionada con las habilidades más complejas.

La codificación consistió en la clasificación de las respuestas³² de los estudiantes de acuerdo con la diversidad de estrategias empleadas para resolver las preguntas. Dicha clasificación se llevó a cabo a partir de un conjunto de criterios específicos definidos previamente para cada pregunta y que figuraban en los manuales de codificación. Estos criterios buscan evidencias del grado de desarrollo de las habilidades evaluadas en las respuestas de los estudiantes.

29. Los docentes que desempeñaron el papel de examinadores en el nivel de su especialidad no pertenecían a las IE de la muestra y fueron capacitados rigurosamente para la aplicación de los instrumentos de manera estandarizada. Cabe señalar que dada la corta edad y características de los estudiantes de segundo grado de primaria, las pruebas fueron aplicadas en grupos de no más de quince estudiantes, de modo que los examinadores pudieran aplicar de manera estandarizada los instrumentos y a la vez establecer condiciones adecuadas para que los estudiantes pudieran resolver la prueba.

30. En promedio los cuadernillos de las pruebas tenían veinte preguntas.

31. Muchos de estos docentes codificadores cuentan con la experiencia de haber participado en anteriores procesos de codificación durante la fase piloto e, incluso, en anteriores evaluaciones nacionales.

32. Nos referimos a todo el proceso mostrado por el estudiante —procedimiento y respuesta— y no solo a la respuesta final.

En efecto, al estar interesados en evaluar el grado de desarrollo de las capacidades evaluadas, se aplicó un modelo de créditos parciales, como se describió en el primer capítulo. Por ejemplo, la respuesta de un estudiante podía ser considerada como totalmente incorrecta y recibir cero puntos, parcialmente correcta y recibir un punto, o correcta y recibir dos puntos. Si bien esta estructura de calificación se basa en tres puntajes posibles, algunos de los ítemes de la EN 2004 fueron calificados con escalas de cuatro o cinco calificaciones posibles.

De otro lado, acompañaba al puntaje un segundo código —de naturaleza cualitativa— que servía para caracterizar el tipo de errores que presentaban algunas de las respuestas de los estudiantes. Cabe precisar que este código no ha influido en la construcción de la escala Rasch.

Para asegurar que el proceso fuera lo más objetivo posible y que todas las respuestas fueran codificadas bajo los mismos criterios, además de trabajar con docentes capacitados y manuales de codificación detallados, se aplicaron mecanismos de control de calidad para evaluar la adecuada aplicación de los criterios por parte de cada codificador a lo largo de los dos meses que duró la codificación.³³

2.10. ¿Cómo se procesaron y analizaron los datos?

Codificadas las pruebas de todos los estudiantes evaluados, se consolidaron y depuraron las bases con los datos, a partir de las cuales se realizaron los análisis psicométricos utilizando la Teoría de Respuesta al Ítem (TRI), en particular el modelo Rasch.³⁴ Como se ha señalado, dicho modelo permite estimar tanto la habilidad de los estudiantes como la dificultad de los ítemes. Además, se realizó un análisis pedagógico que consistió en el estudio del comportamiento de cada uno de los ítemes y de sus criterios de codificación, de manera que la prueba en su conjunto reflejara fielmente el enfoque formulado en el marco de trabajo de cada una de las áreas. Una idea que se debe resaltar es la búsqueda de una evaluación que permita recoger información de un aprendizaje funcional, formativo y útil a la vez, que trascienda los muros de la escuela y que se refleje en la mejora de la calidad de vida de los ciudadanos.

Por otro lado, los cuestionarios de factores asociados fueron procesados y analizados de muy diversas maneras debido a su distinta naturaleza y formato. Así, algunos cuestionarios han sido analizados mediante la TRI y otros tipos de análisis —como el análisis multivariado— para identificar factores que están relacionados con las variables investigadas.

2.11. ¿Cómo se establecieron los niveles de desempeño para las pruebas de rendimiento?

Como señalamos líneas arriba, el modelo de evaluación de la EN 2004 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.

33. Para más detalle ver Anexo 3: Proceso de corrección múltiple de los ítemes abiertos.

34. Este modelo ha sido descrito ampliamente en el capítulo anterior.

La UMC realizó varias consultas a grupos de expertos en educación³⁵ con la finalidad de determinar cuáles son los niveles de desempeño pertinentes para clasificar a los estudiantes según su desempeño en las pruebas. 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 formando una escala de habilidad en la que se determinaron tres niveles de desempeño: suficiente, básico y previo —ordenados de mayor a menor dificultad—. Los expertos fueron convocados para que establecieran los límites que indicaran 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. A este procedimiento se le conoce como establecimiento de puntos de corte.³⁶

De esta manera, la población evaluada ha podido ser categorizada en función de los niveles de desempeño definidos para la prueba y se ha obtenido el porcentaje de población que pertenece a cada uno de dichos niveles.

Se presenta a continuación la descripción general de cada uno de los niveles establecidos a partir de la EN 2004:

Nivel Suficiente: Los estudiantes ubicados en este nivel demuestran un manejo suficiente y necesario de las capacidades evaluadas en el grado. Esto quiere decir que demuestran las habilidades que se esperan de un estudiante que está por terminar el grado. Cabe mencionar que no son estudiantes avanzados ni destacados los que predominan en este nivel, sino adecuados para el grado. Por ello, al finalizar el grado, todos o la gran mayoría de los estudiantes deberían encontrarse en este nivel.

Nivel Básico: Los estudiantes agrupados en este nivel demuestran un dominio incipiente o un manejo elemental de las capacidades evaluadas en el grado. Esto quiere decir que a pesar de que ya van a concluir el grado, solo demuestran un desarrollo parcial de las habilidades esperadas.

Nivel Previo: Los estudiantes en este nivel demuestran un manejo de las capacidades correspondiente a grados anteriores. Esto quiere decir que, a pesar de estar por concluir el grado, los estudiantes de este nivel no han logrado evidenciar el dominio de las habilidades que se esperan de un estudiante que está empezando el grado.

Estos son los tres niveles de desempeño definidos para la EN 2004; sin embargo, debido que hay estudiantes que no lograron responder todo el conjunto de preguntas necesarias para ser considerados en el nivel previo, se ha tenido que definir un grupo llamado «por debajo del previo»:

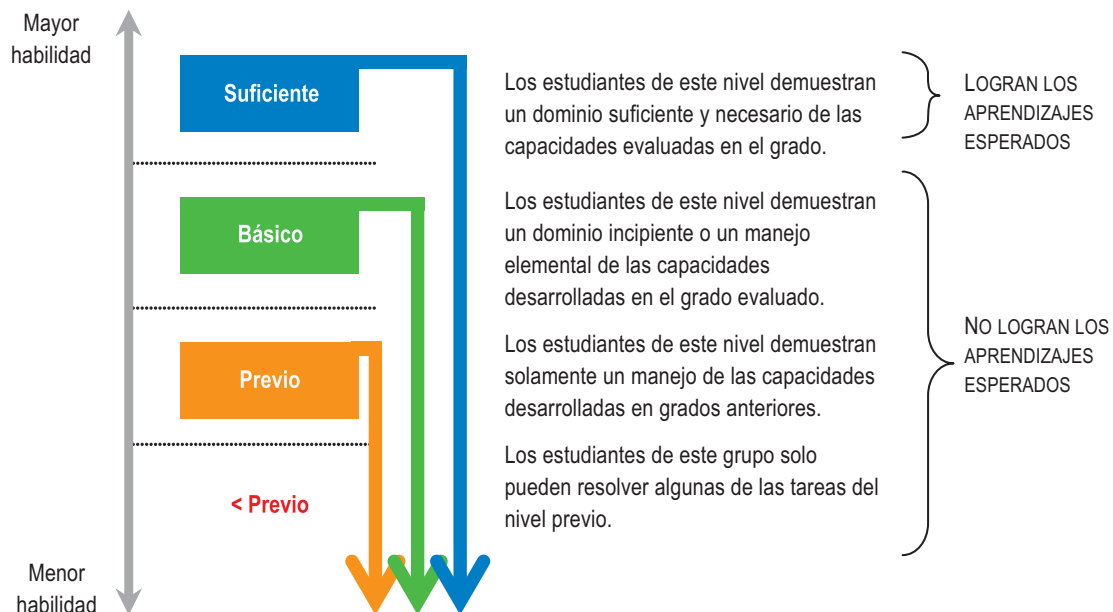
Por Debajo del Previo (< Previo): en este grupo, se ubican los estudiantes que no logran resolver todas las tareas del nivel previo: solo resuelven algunas de ellas, las más sencillas, que generalmente corresponden a ciclos anteriores.

35. Se consultaron alrededor de 160 personas, entre docentes, representantes de editoriales, investigadores, especialistas del área pedagógica de las instancias de gestión descentralizadas y del Ministerio de Educación, etc. para que determinaran los puntos de corte entre los niveles. Es decir, se convocó no solo a expertos teóricos, sino también a expertos prácticos, conocedores de la realidad, de los intereses y de las necesidades de los estudiantes de diversas zonas del país.

36. Para mayor información, véase el Anexo 4: Establecimiento de puntos de corte.

Es importante precisar que estos niveles son inclusivos. Es decir, los estudiantes que se encuentran en el nivel básico pueden resolver preguntas que pertenecen a ese nivel y al nivel previo. Del mismo modo, los estudiantes que están en el nivel suficiente pueden resolver todas las preguntas de los niveles suficiente, básico y previo.

Gráfico 2.2 Niveles de desempeño de la EN 2004



2.12. Descripción de los niveles de desempeño por áreas y grados

Con el fin de comprender mejor los niveles de desempeño, a continuación se presenta la descripción de las habilidades de los estudiantes ubicados en cada uno de estos. Adicionalmente, en el caso del nivel suficiente, la descripción se ilustrará con algunos ejemplos de preguntas que los estudiantes ubicados en él pueden resolver con éxito, así como los puntajes Rasch asociados a dichas tareas.³⁷

37. En este capítulo solo se presentan algunos ejemplos que ilustran las tareas que pueden hacer los estudiantes ubicados en el nivel suficiente. Para ver ejemplos de ítems de este nivel y de los otros niveles de desempeño, puede consultar los informes pedagógicos de resultados de la EN 2004 (UMC 2005).

Cuadro 2.2. Niveles de desempeño en Comprensión de textos escritos: Segundo grado de primaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en este nivel son capaces de: • Localizar información o datos explícitos en el texto, solicitados de forma literal en las preguntas y también de manera parafraseada. • Localizar información o datos explícitos ubicados en partes muy notorias del texto, como en el título o en las primeras líneas del texto, o en partes no tan evidentes, como en medio de un párrafo. • Identificar secuencias temporales en textos de más de tres acciones o sucesos (reconocer qué sucedió antes o después de un evento). • Realizar inferencias sencillas a partir, sobre todo, de información explícita del texto (las inferencias exigen muy poca apelación a información extratextual o al conjunto de saberes con los que cuenta el estudiante). • Relacionar ideas explícitas para realizar inferencias sobre todo locales (las inferencias que exigen lecturas globales son menos frecuentes). • Reconocer relaciones de causa-efecto entre dos ideas que no son sucesivas sino que se hallan distantes (lo que exige realizar una lectura más global del texto). • Reconocer el significado de palabras y expresiones utilizando la información que el texto brinda. • Reconocer el tema central que se presenta de manera general en los textos, cuando este se repite a lo largo del texto. • Identificar el propósito y el receptor de textos a partir de información explícita. • Opinar sobre el contenido del texto (generalmente sobre comportamientos o posiciones que asumen los personajes de historias) y sustentarlo con argumentos que el texto mismo provee o de su conocimiento del mundo más inmediato. En este sentido, la toma de posición implica un grado mínimo de distanciamiento del estudiante respecto del texto para evaluarlo. Asimismo, los estudiantes de este nivel logran leer: • Textos narrativos breves como fábulas, que combinan la narración en tercera persona con intronismos de diálogos (que pueden no presentar dibujos que apoyen el texto verbal). • Textos descriptivos breves, como noticias, cartas, recetas, notas enciclopédicas, cuadros de doble entrada y carteles. Generalmente, son de tema familiar y de vocabulario cotidiano y sencillo. Las notas enciclopédicas desarrollan un único tema y las recetas tienen tres o cuatro pasos en el procedimiento. Los carteles son, por lo general, informativos (presenta un mensaje donde se da a conocer algo), combinan texto verbal e imagen, y, por lo general, presentan una variedad de elementos paratextuales (ilustraciones, tipo y tamaño de letra determinados, formato adecuado, etc.), por lo que su comprensión resulta más sencilla. • Textos cuyos usos son recreativos (textos de ficción como cuentos y fábulas), educacionales (artículos enciclopédicos) y públicos (las noticias o los carteles), aunque estos últimos en menor proporción que los dos primeros.	Los estudiantes ubicados en este nivel son capaces de: • Localizar datos o información explícitos en el texto si son solicitados de forma literal en las preguntas (con las mismas palabras que aparecen en el texto). • Localizar datos o información explícitos en el texto cuando se encuentran en partes muy notorias de los textos, como en el título o en las primeras líneas, expresados en una frase u oración. Esta información, generalmente, responde a las preguntas ¿qué?, ¿quién?, ¿quiénes? • Identificar secuencias temporales en cuentos cortos de tres núcleos narrativos básicos (reconocer qué sucedió al final o al comienzo de un cuento) • Relacionar dos ideas explícitas en textos breves para realizar inferencias locales, fragmentarias y directas. • Reconocer relaciones de causa-efecto entre dos ideas explícitas y sucesivas. • Deducir sentimientos o estados de ánimo de personajes a partir de sus acciones en la historia. • Reconocer tipos de textos por la forma o silueta. Asimismo, los estudiantes de este nivel logran leer: • Textos narrativos, como cuentos breves y sencillos, generalmente compuestos por núcleos narrativos básicos y bien definidos (inicio, conflicto y desenlace). Estos cuentos se encuentran apoyados por dibujos (en forma de viñetas) y son narrados en tercera persona, sin inclusión de diálogos. • Textos descriptivos muy breves, como notas, cartas y carteles. Generalmente, son de tema familiar, y de vocabulario cotidiano y sencillo. Los artículos enciclopédicos desarrollan un único tema. Los carteles tienen muy poca información escrita y pueden incluir imágenes o dibujos. • Textos cuyos usos son recreativos (fábulas) y educacionales (notas enciclopédicas).	Los estudiantes ubicados en este nivel son capaces de: • Identificar el referente gráfico de palabras y oraciones (simples y compuestas), es decir, reconocer el dibujo que representa una palabra u oración. • Identificar palabras (sustantivos, verbos) que son necesarias para completar el sentido de una oración. • Identificar información explícita y literal a partir de la lectura de oraciones sueltas. En este nivel, los estudiantes realizan tareas que están relacionadas, principalmente, con la comprensión de oraciones aisladas y no de textos completos.

Cuadro 2.3. Ejemplos de preguntas de Comprensión de textos escritos: Segundo grado de primaria

Ejemplos de preguntas									
<table><tr><th>Niveles</th><th>Rango de dificultad</th></tr><tr><td>Suficiente</td><td>Entre 380 y 304</td></tr><tr><td>Básico</td><td>Entre 303 y 259</td></tr><tr><td>Previo</td><td>258 o menos</td></tr></table>	Niveles	Rango de dificultad	Suficiente	Entre 380 y 304	Básico	Entre 303 y 259	Previo	258 o menos	Se les presentó a los estudiantes un texto descriptivo breve que desarrolla como tema el origen y la utilidad del fuego. Para responder a este ítem, los estudiantes debían establecer relaciones entre las partes del texto para inferir aquella idea que alude de manera general al tema central e identificarla entre tres opciones posibles. En esta tarea los estudiantes deben superar un nivel de comprensión local y fragmentaria, y aproximarse a una comprensión global del texto. Se les mostró a los estudiantes un aviso escrito por una niña en el que se describe brevemente las características de una casaca extraviada y se señala dónde debe ser llevada en caso de ser encontrada. Para responder este ítem, los estudiantes debían deducir el propósito del texto entre tres opciones propuestas. Se les presentó a los estudiantes un texto narrativo, narrado en tercera persona y sin inclusión de diálogos, en el que intervienen dos personajes: un perro y un pollo. En la historia, el perro se burla del pollo ofreciéndole algo que este no come. Posteriormente, el pollo hace lo mismo en venganza. Para responder este ítem, los estudiantes debían deducir cuál es la razón por la que el pollo en la historia no come nada de lo que el perro le ofrece. Así, debía establecer como causa el hecho de que la comida que el perro le da (huesos) no es una comida que un pollo normalmente come. Así, este ítem demanda un mínimo de conocimientos externos al texto sobre lo que comen típicamente un perro y un pollo.
Niveles	Rango de dificultad								
Suficiente	Entre 380 y 304								
Básico	Entre 303 y 259								
Previo	258 o menos								
	La lectura nos informa principalmente sobre: a) los alimentos crudos. ✓ b) el fuego. c) los incendios. Lo que has leído: ✓ a) nos pide que ayudemos a encontrar una casaca. b) nos recomienda comprar casacas como la de Rosa. c) nos aconseja que no debemos olvidar nuestras casacas. ¿Por qué el pollo no comió nada? a) Porque no tenía hambre. b) Porque estaba molesto con el perro. ✓ c) Porque el pollo no come huesos.								

Cuadro 2.4. Niveles de desempeño en Comprensión de textos escritos: Sexto grado de primaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en este nivel son capaces de: • Localizar datos explícitos en el texto. Estos datos son solicitados de forma literal en las preguntas (con las mismas palabras con que aparecen en el texto) y también de manera parafraseada (con otras palabras). • Localizar datos o información explícitos ubicados en partes muy notorias del texto, como en el título o en las primeras líneas del texto, o pueden encontrarse en partes no tan evidentes, como en medio de un párrafo. • Interpretar y establecer conexiones globales en el texto. Así, pueden hacer deducciones que van más allá de la información explícita del texto: inferen, por ejemplo, la idea principal del texto y la intención comunicativa del autor. • Reconocer relaciones entre dos ideas explícitas e implícitas que no son sucesivas sino que se hallan más distantes, lo que exige realizar una lectura global del texto. • Reconocer relaciones causales, de comparación, de contraste. • Identificar a qué se refiere una expresión que sustituye a otra. • Reconocer el significado de palabras utilizando información que el texto brinda. • Deducir el receptor al que se dirigen los textos cuando se encuentra implícito. • Asumir un punto de vista sobre el contenido del texto y sustentarlo con argumentos que extraen principalmente del texto o de su mundo personal. • Evaluar algunos recursos formales: comillas, paréntesis, y dibujos o imágenes que usa el autor para transmitir significado. Asimismo, los estudiantes de este nivel logran comprender globalmente: • Textos narrativos, como cuentos y leyendas, en los que aparecen varios personajes y suceden diversas acciones. Pueden ser narrados en primera o tercera persona (la voz de los personajes se introduce con guiones o con comillas). • Textos expositivos, como artículos enciclopédicos de temas variados. Generalmente, contienen temas familiares y su lenguaje es sencillo y formal. La estructura es evidente pues cada subtema está claramente desarrollado en cada párrafo. Sin embargo, los textos son relativamente extensos. • Textos descriptivos, como noticias y cuadros estadísticos de variada información. • Textos argumentativos, como pequeños editoriales de diarios o revistas, cuyo lenguaje es sencillo e informal y de temática familiar. Muestran una estructura libre y se ilustra la opinión con casos y ejemplos concretos de la vida cotidiana. Asimismo, se encuentran afiches publicitarios con escasa información verbal. La temática de estos afiches gira en torno de campañas en bien de la comunidad. • Textos cuyos usos de lectura son recreativos (textos de ficción, como cuentos), educacionales (textos diseñados para enseñar o educar, como artículos enciclopédicos), y públicos. Estas lecturas de contexto público tratan temas muy relacionados con el entorno social del estudiante, tales como la conservación del medio ambiente y las campañas de salud. Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.	Los estudiantes ubicados en este nivel son capaces de: • Localizar datos o información explícitos en el texto si son solicitados de forma literal en las preguntas (con las mismas palabras que aparecen en el texto). • Localizar datos o información explícitos en el texto cuando se encuentran en partes muy notorias de los textos, como en el título o en las primeras líneas, expresados en una frase u oración. Esta información, generalmente, responde a las preguntas ¿qué?, ¿quién?, ¿quiénes? • Identificar secuencias temporales en cuentos cortos de tres núcleos narrativos básicos (reconocer qué sucedió al final o al comienzo de un cuento) • Relacionar dos ideas explícitas en textos breves para realizar inferencias locales, fragmentarias y directas. • Reconocer relaciones de causa-efecto entre dos ideas explícitas y sucesivas. • Deducir sentimientos o estados de ánimo de personajes a partir de sus acciones en la historia. • Reconocer tipos de textos por la forma o silueta. Asimismo, los estudiantes de este nivel logran leer: • Textos narrativos, como cuentos breves y sencillos, generalmente compuestos por núcleos narrativos básicos y bien definidos (inicio, conflicto y desenlace). Estos cuentos se encuentran apoyados por dibujos (en forma de viñetas) y son narrados en tercera persona, sin inclusión de diálogos. • Textos descriptivos muy breves, como notas, cartas y carteles. Generalmente, son de tema familiar, y de vocabulario cotidiano y sencillo. Los artículos enciclopédicos desarrollan un único tema. Los carteles tienen muy poca información escrita y pueden incluir imágenes o dibujos. • Textos cuyos usos son recreativos (fábulas) y educacionales (notas enciclopédicas). Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.	Los estudiantes ubicados en este nivel son capaces de: • Identificar el referente gráfico de palabras y oraciones (simples y compuestas), es decir, reconocer el dibujo que representa una palabra u oración. • Identificar palabras (sustantivos, verbos) que son necesarias para completar el sentido de una oración. • Identificar información explícita y literal a partir de la lectura de oraciones sueltas. En este nivel, los estudiantes realizan tareas que están relacionadas, principalmente, con la comprensión de oraciones aisladas y no de textos completos. Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.

Cuadro 2.5. Ejemplos de preguntas de Comprensión de textos escritos: Sexto grado de primaria

Ejemplos de preguntas									
<table border="1"> <tr> <th>Niveles</th><th>Rango de dificultad</th></tr> <tr> <td>Suficiente</td><td>Entre 386 y 277</td></tr> <tr> <td>Básico</td><td>Entre 276 y 257</td></tr> <tr> <td>Previo</td><td>256 o menos</td></tr> </table>	Niveles	Rango de dificultad	Suficiente	Entre 386 y 277	Básico	Entre 276 y 257	Previo	256 o menos	Se les mostró a los estudiantes un afiche cuya finalidad es persuasiva: su intención es lograr la modificación de una conducta que, en este caso, está relacionada con el cuidado del peso. Así, se censuran ciertos hábitos alimenticios por ser considerados nocivos para la salud. El texto presenta una imagen bastante expresiva y sugerente con escasa información verbal. Los estudiantes debían inferir la intención comunicativa del autor del afiche: persuadir a la gente de conservar la salud descartando hábitos alimenticios negativos. Para contestar esta pregunta, los estudiantes requieren de habilidades inferenciales que les permitan entender las intenciones planteadas en cada una de las alternativas. Se les presentó a los estudiantes un texto argumentativo breve en el que se sustenta claramente un punto de vista a favor de un tema sobre el que existen opiniones diversas: la crianza de animales domésticos en los hogares. En ese sentido, el texto pretende persuadir al lector con el fin de modificar cualquier opinión contraria a la crianza de dichos animales. Los estudiantes debían reconocer la relación entre dos párrafos iniciales del texto y discriminar argumentos sencillos y opuestos. Para contestar esta pregunta, los estudiantes requieren de habilidades inferenciales y comprensión. Se les mostró a los estudiantes un texto narrativo que relata un hecho fantástico con cierto tono humorístico. El narrador —que también es personaje— cuenta en primera persona un hecho que le tocó vivir en circunstancias muy singulares. A pesar de no mencionar explícitamente qué o quién es, el lector puede inferir que es un insecto el que narra dicha aventura en la cabeza de un calvo. Es un texto extenso, de estructura libre y con un lenguaje sencillo. Los estudiantes debían establecer con claridad las relaciones de causa-efecto planteadas en el texto. Para contestar esta pregunta, los estudiantes debían poner en juego sus habilidades inferenciales de causalidad desde la comprensión global del texto narrativo.
Niveles	Rango de dificultad								
Suficiente	Entre 386 y 277								
Básico	Entre 276 y 257								
Previo	256 o menos								
362 → 310 → 303 →	Con este afiche se busca principalmente: a) informar sobre las enfermedades del corazón. b) describir en qué consiste la obesidad. ✓ c) aconsejar al público que cuide su peso. d) ordenar a los gordos que adelgace «Existen muchas personas que afirman que no hay nada más terrible que vivir con mascotas porque estas lo único que hacen es ensuciar y, en el caso de perros y gatos, hacer destrozos por toda la casa». «Asimismo, científicamente se ha demostrado que tener una mascota —perro, gato, canario...— ayuda a las personas a no estar tan tensas y a relacionarse mejor con los demás». Los dos párrafos anteriores indican: ✓ a) opiniones diferentes sobre las mascotas. b) opiniones a favor de las mascotas. c) opiniones en contra de las mascotas. d) las mismas opiniones sobre las mascotas. ¿Por qué el personaje que vive en la cabeza del calvo no quiere que el último pelo de la cabeza sea arrancado?								

Cuadro 2.6. Niveles de desempeño en Comprensión de textos escritos: Tercer grado de secundaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes de este nivel son capaces de: • Localizar datos explícitos en el texto: estos pueden estar tanto en partes evidentes (las primeras líneas, el final o el título) como en partes poco notorias insertadas en el cuerpo del texto. • Ubicar determinada información dentro de un grupo de datos semejantes al dato solicitado (en el texto o en la pregunta). • Hacer inferencias que involucren una comprensión global del texto. • Deducir y seleccionar la idea principal de un texto entre una serie de alternativas. • Interpretar e integrar ideas contenidas en el texto, muchas veces distantes entre sí. • Interpretar relaciones semánticas de causalidad, equivalencia y contraste entre ideas del texto. • Interpretar metáforas y analogías. • Reconocer a qué palabra o expresión se refiere el pronombre enclítico (como <i>lo, la, le</i> insertos en el verbo) y los pronombres que aluden a eventos (<i>ello, esto</i>). • Interpretar el significado de palabras poco conocidas en su entorno escolar, lo que involucra comprender el texto en su conjunto y/o tener una mayor experiencia lectora. • Interpretar el mensaje de textos ícono-verbales, como afiches o carteles. • Reconocer el propósito del autor de un texto que presenta dos o más opiniones diversas sobre un tema determinado. • Deducir a quién se dirige el texto, aun cuando el receptor se encuentra implícito. • Opinar sobre actitudes y acciones de personajes en textos narrativos y relacionar su punto de vista con su vida personal y cotidiana. • Sustentar su opinión sobre los argumentos de un texto. • Evaluar los recursos formales (tipográficos, elementos gráficos) que usa el autor en el texto para transmitir el sentido. • Reconocer la función de signos convencionales que están fuera del cuerpo principal del texto, como las llamadas a pie de página. Asimismo, los estudiantes de este nivel logran leer globalmente: • Textos narrativos de carácter literario en los que se construye mundos ficcionales y cuyos finales son inesperados. • Textos descriptivos breves, como artículos de revistas, diarios o enciclopedias. Estos presentan una estructura evidente (con títulos, subtítulos o párrafos por cada subtema). Su temática es científica, social o cultural, y contiene un lenguaje claro y directo. Asimismo, incluyen cuadros, tablas, diagramas de barras. • Textos expositivos de estructura evidente (con títulos y subtítulos) que alternan un vocabulario formal con explicaciones dadas por especialistas mediante discursos indirectos. • Textos argumentativos breves de estructura evidente, como artículos de revistas, cartas y textos de opinión múltiple; y afiches publicitarios, que intentan convencer al lector de comprar, elegir un producto, utilizar un servicio, etc. • Textos recreativos (cuentos) y educacionales (textos de material educativo), y, de manera incipiente, textos de uso público frecuentes en la sociedad adulta. Los textos anteriores pueden contener tanto información verbal como ícono-verbal.	Los estudiantes de este nivel son capaces de: • Ubicar datos explícitos que se encuentran en partes poco notorias del texto. • Interpretar e integrar ideas sucesivas y cercanas entre sí. • Deducir y seleccionar la idea principal de un texto entre una serie de alternativas. • Interpretar relaciones semánticas entre ideas, que pueden ser de causalidad, equivalencia, contraste, etc. • Reconocer a qué se refiere un pronombre personal cuando sustituye a personas o sustantivos concretos. • Reconocer el significado de palabras sencillas y proponer un término sinónimo. • Deducir el propósito del autor en un texto. • Deducir a quién está dirigido el texto, aun cuando el receptor se encuentre implícito. • Opinar sobre actitudes de personajes en textos narrativos. • Evaluar la estructura o disposición de la información en los textos. Asimismo, los estudiantes de este nivel logran leer globalmente: • Textos narrativos sencillos de carácter literario. • Cuadros, tablas, diagramas de barras y otros textos descriptivos sencillos. • Textos descriptivos breves, como artículos de revistas o de diarios, y crónicas periodísticas. En su vocabulario, predomina un lenguaje familiar y su estructura es de párrafos cortos. • Textos expositivos breves cuya estructura es evidente, por cada párrafo, hay un subtema claramente diferenciado. Su lenguaje es formal y sencillo. • Textos argumentativos, como textos de opinión breves. Generalmente, contienen un vocabulario sencillo y cotidiano, y su información está distribuida en párrafos o columnas. • Afiches publicitarios. Los textos anteriores pueden contener tanto información verbal como ícono-verbal.	Los estudiantes de este nivel son capaces de: • Ubicar uno o más datos explícitos que aparecen en el texto. Estos datos se encuentran en partes del texto. • Ubicar datos que se distinguen claramente de los demás datos presentados en el texto o en la pregunta. • Hacer inferencias locales interpretando un fragmento específico del texto. Deducir y seleccionar la idea principal de un texto cuando esta es muy obvia. • Interpretar relaciones semánticas sencillas y evidentes entre ideas explícitas de un texto. • Reconocer a qué se refiere un pronombre personal que alude a personas u objetos. • Deducir el propósito del autor. • Deducir a quién está dirigido el texto cuando el receptor es bastante evidente. • Opinar sobre actitudes de personajes en textos narrativos. • Justificar su postura con argumentos obvios o esperados extraídos del mismo texto. Asimismo, los estudiantes de este nivel logran leer globalmente: • Textos narrativos breves, como cuentos o leyendas de corta extensión, cuyos argumentos, personajes y espacios les resultan bastante familiares. • Textos expositivos breves de estructura evidente y temas cotidianos con lenguaje familiar e, incluso, coloquial. • Textos descriptivos breves, sencillos y de estructura evidente (con títulos, subtítulos) sobre objetos y lugares, como pequeñas crónicas, noticias breves, artículos de diarios. • Textos argumentativos de opinión breves. • Textos cuyo uso es recreativo y personal (cuentos), y educacional (textos de material educativo). En este sentido, leen textos que hacen alusión a temas relacionados con un universo adolescente y escolar. Los textos anteriores pueden contener tanto información verbal como ícono-verbal.

Cuadro 2.7. Ejemplos de preguntas de Comprensión de textos escritos: Tercer grado de secundaria

Ejemplos de preguntas									
<table><tr><th>Niveles</th><th>Puntajes</th></tr><tr><td>Suficiente</td><td>Entre 548 y 500</td></tr><tr><td>Básico</td><td>Entre 499 y 456</td></tr><tr><td>Previo</td><td>456 o menos</td></tr></table>	Niveles	Puntajes	Suficiente	Entre 548 y 500	Básico	Entre 499 y 456	Previo	456 o menos	El texto narrativo «La historia de lasá» es una leyenda de la tradición oral de un pueblo indígena de la selva de nuestro país y tiene como propósito explicar cómo se formó el arcoíris. La historia es bastante simple y repite la estructura de muchas historias conocidas por los estudiantes: la relación amorosa de dos jóvenes es interrumpida por un personaje perverso que desea separarlos. En esta pregunta, los estudiantes debían identificar la función de un recurso formal —del asterisco— en el texto: la explicación del término <i>cashinahua</i> ubicada en una nota a pie de página. Para contestar este ítem, los estudiantes requieren de habilidades reflexivas para evaluar recursos formales muy frecuentes. Se les mostró a los estudiantes un texto periodístico en el que se presentan, en columnas, las opiniones de cinco personas entrevistadas en la calle acerca de cómo disminuir la delincuencia en la ciudad. El lenguaje es sencillo y sumamente coloquial. Los estudiantes debían deducir la idea que tienen en común las opiniones de dos personas. Para esto, debe ubicar los testimonios correspondientes y luego compararlos e inferir cuál es la idea que se repite en los dos («hacen falta más policías»). Para contestar este ítem los estudiantes deben poner en juego sus habilidades inferenciales. Se les presentó a los estudiantes el texto argumentativo «Perú, país multilingüe». Este texto expone algunas razones que sustentan la necesidad de preservar las lenguas nativas en el Perú y, por ello, demanda medidas a los gobernantes para que fomenten su conservación. Los estudiantes debían reflexionar sobre el enunciado que considera a las lenguas nativas como un obstáculo para el progreso social y emitir una opinión que puede basarse en los argumentos del texto. Para contestar esta pregunta, los estudiantes deben poner en juego sus habilidades reflexivas.
Niveles	Puntajes								
Suficiente	Entre 548 y 500								
Básico	Entre 499 y 456								
Previo	456 o menos								
	En el comienzo del texto aparece un asterisco (*) al lado de la palabra <i>cashinahua</i>. ¿Para qué crees que el autor utilizó este signo?								
	¿Qué idea tienen en común Armando Guevara y Ruth Magan?								
	Algunas personas opinan que, en el Perú, lenguas como el quechua y el aimara son una barrera para el progreso del país. ¿Estás de acuerdo con estas personas? Explica tu respuesta tomando en cuenta —bien sea para apoyarlas o para rechazarlas— las ideas del texto—“Perú, país multilingüe”.								

Cuadro 2.8. Niveles de desempeño en Comprensión de textos escritos: Quinto grado de secundaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes de este nivel son capaces de: • Ubicar uno o más datos explícitos que aparecen en el texto. Estos datos se encuentran en partes notorias, como al inicio o fin del texto, o en otras partes insertadas en el cuerpo del texto. • Relacionar varios datos lejanos entre sí para ubicar la información. • Hacer inferencias que involucren una comprensión global del texto. • Elaborar sintéticamente la idea principal del texto. • Interpretar e integrar ideas contenidas en el texto, muchas veces distantes entre sí, inesperadas o contrarias a las expectativas del estudiante. • Interpretar relaciones semánticas entre ideas que pueden ser de causalidad, consecuencia, equivalencia, contraste, etc. • Interpretar y relacionar una serie de ideas en las que puede encontrarse semejanzas, comparaciones, metáforas, correspondencias, analogías. • Reconocer a qué palabra o expresión se refiere un pronombre enclítico (como <i>lo, la, le</i> insertos en el verbo) o pronombres que aluden a eventos (<i>ello, esto</i>). • Interpretar el significado de palabras poco conocidas en su entorno escolar, lo que involucra comprender el texto en su conjunto y/o una mayor experiencia lectora. • Esta tarea la realiza considerando solamente el contexto de la lectura, sin acudir al diccionario. • Deducir el propósito del autor de un texto que presenta dos o más opiniones diversas sobre un tema determinado. • Deducir a quién se dirige el texto, aun cuando el receptor se encuentre implícito. • Opinar sobre actitudes de personajes en textos narrativos. • Sustentar su opinión sobre los argumentos de un texto y evaluar su calidad persuasiva. • Relacionar su punto de vista acerca de lo leído con casos concretos de la vida real. • Evaluar los recursos formales (discursos formales y coloquiales) que usa el autor en el texto para transmitir el sentido.	Los estudiantes de este nivel son capaces de: • Ubicar uno o más datos explícitos que aparecen en el texto. Estos datos se encuentran en partes notorias, como al inicio o al final del texto, o insertadas en el cuerpo del texto. • Ubicar la información entre un grupo de datos semejantes (en el texto o en la pregunta) al dato solicitado. • Interpretar e integrar ideas sucesivas y cercanas entre sí. • Deducir y seleccionar la idea principal de un texto entre una serie de alternativas que compiten entre sí. • Interpretar relaciones semánticas entre ideas que pueden ser de causalidad, equivalencia, contraste, etc. • Interpretar y relacionar una serie de ideas en las que pueden encontrarse comparaciones y analogías. • Reconocer a qué se refiere un pronombre personal cuando sustituye a sustantivos abstractos. • Reconocer el significado de palabras sencillas y proponer un término sinónimo. Esta tarea la realiza el estudiante considerando solamente el contexto de la lectura, sin acudir al diccionario. • Deducir el propósito del autor de un texto cuyo tema es evidente. • Deducir a quién está dirigido el texto aun cuando el receptor se encuentre implícito. • Opinar sobre actitudes de personajes en textos narrativos. • Reconocer las razones de textos argumentativos y/o evaluar su calidad persuasiva. • Evaluar la estructura o disposición de la información en los textos. Asimismo, los estudiantes del nivel básico logran comprender globalmente: • Textos narrativos breves (cuentos, adaptaciones) de carácter literario con una estructura básica. • Textos expositivos breves de estructura evidente y no evidente. Generalmente, son de naturaleza social y cultural, como artículos de revistas o diarios, infografías, etc. En su vocabulario, predomina un lenguaje familiar.	Los estudiantes de este nivel son capaces de: • Ubicar uno o más datos explícitos que aparecen en el texto. Estos datos se encuentran en partes notorias, como al inicio o al final del texto, o insertadas en el cuerpo del texto. • Ubicar datos que se distinguen claramente de los otros datos presentados en el texto o en la pregunta. • Hacer inferencias locales interpretando un fragmento específico del texto. • Deducir y seleccionar la idea principal de un texto entre una serie de alternativas. • Interpretar relaciones semánticas sencillas y evidentes entre ideas explícitas de un texto. • Interpretar relaciones de causalidad y comparaciones. • Reconocer el significado de términos sencillos o que puede ser deducido por prefijos o sufijos. • Deducir el propósito del autor de un texto que presenta un tema evidente. • Deducir a quién está dirigido el texto cuando el receptor es bastante explícito o evidente. • Opinar sobre actitudes de personajes en textos narrativos. • Justificar su postura con argumentos obvios o esperados extraídos del mismo texto. • Identificar los recursos formales que usa el autor en un texto para transmitir el sentido. Asimismo, los estudiantes del nivel previo logran comprender globalmente: • Textos narrativos breves cuyos personajes y ambientes son bastante juveniles, como leyendas. • Textos expositivos breves de estructura evidente con subtítulos. Generalmente, sus temas forman parte de la cotidianidad social, como artículos de revistas o diarios. En su vocabulario, predomina un lenguaje familiar e, incluso, coloquial.

Cont. Cuadro 2.8

Asimismo, los estudiantes de este nivel logran comprender globalmente:

- Textos narrativos de carácter literario (cuentos) en los que se trata con mundos ficcionales complejos de finales inesperados. Pueden ser historias paralelas o absurdas, con cierto lenguaje irónico y ambiguo.
- Textos expositivos extensos de estructura evidente (con títulos, subtítulos, párrafos subtemáticos) y no evidente (sin señales explícitas de subtemas). Generalmente, son de naturaleza científica o social, como artículos de revistas o diarios, infografías, etc. Su vocabulario alterna un lenguaje formal y técnico con uno familiar y cotidiano. Frecuentemente, son textos que contienen citas de especialistas en el tema mediante el discurso directo.
- Textos descriptivos con variada información sobre objetos y lugares (cuadros estadísticos, infografías), así como crónicas periodísticas. El lenguaje es formal a pesar de que inserta algunos comentarios coloquiales de testigos de los hechos.
- Textos argumentativos breves y extensos de estructura evidente o no tan evidente.
Generalmente, son escritos académicos de vocabulario formal y técnico, como afiches publicitarios que pretenden convencer de la compra de un producto o de la calidad de un servicio.
- Textos cuyo uso es recreativo y personal (cuentos), educacional (textos de material educativo) y público (textos frecuentes en la sociedad adulta). En este sentido, leen textos cuyos temas están relacionados con un universo cada vez más adulto y con el que los estudiantes están próximos a insertarse.

Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.

- Textos descriptivos breves con variada información sobre objetos y lugares (cuadros estadísticos, certificados), así como crónicas breves y sencillas.
- Textos argumentativos, como textos de opinión breves. Generalmente, contienen un vocabulario sencillo y cotidiano, y su información está distribuida en párrafos o columnas.
- Textos de uso recreativo y personal (cuentos), o educacional (textos de material educativo); y, de manera incipiente, textos cuyo propósito es público (textos frecuentes en la sociedad adulta).

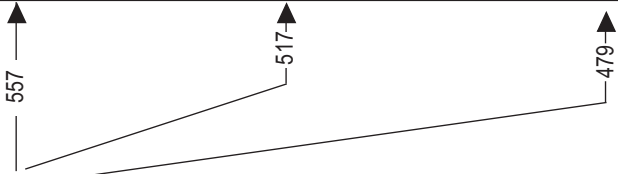
Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.

- Textos descriptivos breves, sencillos y de estructura evidente (con subtítulos) sobre objetos y lugares, como artículos de diarios, cuadros estadísticos.
- Textos argumentativos, como textos de opinión breves que recogen puntos de vista de ciudadanos de la calle. Contiene un vocabulario sencillo e, incluso, coloquial.
- Textos de uso recreativo y personal (cuentos), y educacional (textos de material educativo). En este sentido, leen textos que aluden a temas relacionados con un universo adolescente y escolar, y, de modo muy primario y elemental, con un universo más adulto.

Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como icono-verbal.

Cuadro 2.9. Ejemplos de preguntas de Comprensión de textos escritos: Quinto grado de secundaria

Ejemplos de preguntas									
<table><tr><th>Niveles</th><th>Puntajes</th></tr><tr><td>Suficiente</td><td>Entre 566 y 475</td></tr><tr><td>Básico</td><td>Entre 474 y 451</td></tr><tr><td>Previo</td><td>450 o menos</td></tr></table>	Niveles	Puntajes	Suficiente	Entre 566 y 475	Básico	Entre 474 y 451	Previo	450 o menos	Se les mostró a los estudiantes un artículo argumentativo que pretende convencer de que la televisión es negativa para los niños y jóvenes. La estructura es evidente, aunque el texto es extenso, y contiene un lenguaje formal y académico. Los estudiantes debían reconocer una relación causal en la que la información es, hasta cierto punto, inoperada para el mundo adolescente: que los efectos perjudiciales de la televisión son ocultados porque las industrias del espectáculo dejarían de percibir grandes ganancias. Para contestar esta pregunta, los estudiantes requieren de habilidades inferenciales que les permitan entender la sustentación de cada uno de los argumentos. Se les mostró a los estudiantes una crónica periodística acerca de un temblor ocurrido en un poblado de Cusco y de la situación penosa de sus habitantes. Su extensión es breve y ha sido extraído de un diario de circulación social. Los estudiantes debían reconocer y explicar la relación entre el título y la idea principal del texto o reconocer la intención del autor de llamar la atención del lector con ese título. Para contestar este ítem, los estudiantes requieren de habilidades reflexivas y una comprensión global del texto. Se les presentó a los estudiantes un texto narrativo extenso de final inesperado. Este relata la historia de un hombre que ha descubierto que ha sido enterrado vivo y de tres profanadores de tumbas que intentan robar su cadáver. Los estudiantes debían interpretar el final del cuento haciendo inferencias a partir de una serie de acciones de los personajes a lo largo del relato. Para contestar este ítem, los estudiantes debían poner en juego sus habilidades inferenciales y lograr la comprensión global del texto.
Niveles	Puntajes								
Suficiente	Entre 566 y 475								
Básico	Entre 474 y 451								
Previo	450 o menos								
	Según el texto, ¿por qué se quiere ocultar los efectos perjudiciales de la televisión? ¿Por qué crees que el autor del texto eligió la frase “¡Señor aplaca tu ira!” como título de la noticia? Henry Armstrong murió: a) enterrado vivo. b) de una grave enfermedad. ✓ c) de un fuerte golpe en la cabeza. d) al recibir el impacto de un relámpago.								



Cuadro 2.10. Niveles de desempeño en Matemática: Segundo grado de primaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en el nivel suficiente pueden calcular la suma de dos números de tres cifras cada uno con una o dos transformaciones, en cualquiera de las posiciones, y calcular restas de números de tres cifras con una transformación. Estos estudiantes pueden completar progresiones aritméticas con razones pequeñas y recodificar números de dos cifras de su representación simbólica a su descomposición en decenas y unidades. Respecto de la gestión y administración del entorno, dominan la lateralidad y combinan relaciones espaciales para ubicar objetos. Además, identifican tanto las formas cúbicas presentes en su entorno, como triángulos y cuadrados, en diversas posiciones y contextos. También pueden clasificar objetos de acuerdo con dos atributos. Respecto de la estructura aditiva, esta se encuentra en proceso de desarrollo pues resuelven diversos tipos de problemas aditivos presentados en distintas formas. En general, no dependen de la forma de presentación para resolverlos; es decir, estos estudiantes resuelven de manera adecuada los problemas que se presentan como un texto continuo y aquellos que se proponen como historia.	Estos estudiantes se encuentran en proceso de construcción de la estructura de orden de los números naturales, lo que les permite ordenarlos y completar progresiones aritméticas crecientes y decrecientes. Asimismo, utilizan el algoritmo convencional para calcular sumas que requieren solo una transformación y restas sin transformación (con minuendo de dos cifras y sustraendos de una o dos cifras). Respecto de la comprensión y administración del entorno, pueden identificar ciertos elementos de cuerpos y figuras geométricas, aunque con nombres coloquiales. Identifican triángulos equiláteros e isósceles cuando están en su posición más usual, y cuadrados en diversas posiciones. Además, han incorporado el uso del vocabulario geométrico básico para nombrar polígonos elementales (triángulo, cuadrado, rectángulo). En cuanto a la orientación espacial, estos estudiantes pueden identificar objetos que se encuentran «debajo de» y «encima de» otro objeto. Asimismo, pueden recodificar números de dos cifras de su representación gráfica a la simbólica y en su descomposición en decenas y unidades, por lo que se puede afirmar que su sistema de numeración posicional está en proceso de construcción. Estos estudiantes también pueden clasificar conjuntos de objetos de acuerdo con su cantidad o de acuerdo con dos atributos perceptivos, y completar una seriación gráfica elemental. En cuanto a los problemas aditivos, los estudiantes de este nivel pueden resolver los problemas más sencillos de transformación de estados y de combinación de cantidades. No dependen de la forma en la que se presentan los problemas, pues resuelven tanto aquellos que aparecen como parte de un texto continuo como aquellos que tienen apoyo gráfico.	Estos estudiantes tienen ya afianzado el concepto de número, lo que les permite representar los números mediante diversos soportes. Además, identifican la cantidad con la notación convencional, es decir, reconocen lo que el numeral significa en términos de la cuantificación como atributo de un conjunto de objetos. Los estudiantes ubicados en este nivel pueden realizar conteos de números naturales hasta el 99 y pueden diferenciar el número mayor del número menor. Sin embargo, lo hacen por visualización de una recta numérica o por el reconocimiento de la secuencia de conteo; por ejemplo, saben que 45 es mayor que 32 «porque viene después en la serie» y no por comprensión del sistema de valor posicional. También han incorporado la noción de adición como proceso de reunir, juntar o aumentar. Asimismo, pueden realizar adiciones y sustracciones con números de una cifra y aplican estos conocimientos para calcular sumas de números de dos cifras sin «llevar». Respecto de la comprensión y la administración de su entorno, estos estudiantes han logrado incorporar relaciones espaciales como «delante de», «detrás de», «encima de», «debajo de». Reconocen su propia lateralidad, identifican en su entorno formas geométricas como el cilindro y el triángulo equilátero en su posición más usual. Finalmente, en cuanto a los problemas aditivos de texto, estos estudiantes resuelven problemas que involucran el cálculo de un total conociendo dos cantidades parciales.

Cuadro 2.11. Ejemplos de preguntas de Matemática: Segundo grado de primaria

Ejemplos de preguntas									
Esta pregunta evalúa la capacidad del estudiante para resolver problemas aditivos presentados en forma de historieta. La situación presentada corresponde a la adaptación a este formato de un problema verbal de transformación de estados, donde la incógnita es, precisamente, la cantidad que se transforma. Para resolverla, el estudiante debía interpretar la información presentada en forma gráfica, identificar los datos —en este caso la cantidad inicial de huevos y la cantidad final—, y reconocer que se ha usado una cantidad de huevos que no se conoce y que el problema le solicita hallar. Para responderla el estudiante debía comprender la situación propuesta, elegir la operación adecuada para el caso, ejecutarla y finalmente interpretar el resultado en el contexto de la situación planteada. Esta pregunta evalúa la capacidad del estudiante de identificar objetos de acuerdo a dos atributos señalados, la dificultad de la pregunta estriba en el uso de la proposición negativa para definir una de las características. Para responder el estudiante podía haber dividido en dos subconjuntos, por ejemplo, uno de perros y otro de gatos. El estudiante debía, primero, identificar que se estaba pidiendo un elemento del complemento del conjunto de perros (Toto no es perro) y, luego, en ese subconjunto identificar a aquel que tiene cola negra. Esta pregunta implica, además de la comprensión lectora, la noción de conectores lógicos. Esta pregunta evalúa la capacidad del estudiante para calcular una suma de dos números de tres dígitos, ordenados en forma vertical, en la que, si se hace uso del algoritmo de cálculo convencional, debería realizar transformación de unidades en dos momentos del proceso de cálculo. Otra estrategia que se puede utilizar era percibir que el segundo sumando era una unidad menor que 300, entonces sumar 300 y, luego, restar una unidad al resultado.	358 318 298 <table><tr><th>Niveles</th><th>Puntajes</th></tr><tr><td>Suficiente</td><td>Entre 369 y 293</td></tr><tr><td>Básico</td><td>Entre 292 y 247</td></tr><tr><td>Previo</td><td>246 o menos</td></tr></table>	Niveles	Puntajes	Suficiente	Entre 369 y 293	Básico	Entre 292 y 247	Previo	246 o menos
Niveles	Puntajes								
Suficiente	Entre 369 y 293								
Básico	Entre 292 y 247								
Previo	246 o menos								
Lee y observa con atención:									
Toto tiene cola negra y no es perro. Márcalo con X.									
Escribe en el el resultado de la operación.									
$\begin{array}{r} 657 \\ + 299 \\ \hline \end{array}$									

Niveles	Puntajes
Suficiente	Entre 369 y 293
Básico	Entre 292 y 247
Previo	246 o menos

Cuadro 2.12. Niveles de desempeño en Matemática: Sexto grado de primaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en el nivel suficiente pueden resolver situaciones problemáticas, rutinarias y no rutinarias, que cuentan con algunos datos implícitos y que, por ello, deben ser deducidos. La resolución de dichas situaciones implica la adaptación o elaboración de una estrategia de solución. Estas estrategias involucran el planteamiento de secuencias de hasta tres operaciones aritméticas básicas y, en algunas ocasiones, la conexión de los diferentes contenidos matemáticos. Además, estos estudiantes poseen un conocimiento y manejo adecuado de la estructura del sistema de numeración decimal, de las expresiones fraccionarias y decimales más usuales, en situaciones cotidianas. Entre los principales contenidos y habilidades implicados en la resolución de problemas están la aplicación de operaciones combinadas con y sin signos de agrupación en el ámbito numérico de los números naturales y fraccionarios, el cálculo de los resultados de adiciones o sustracciones con números decimales hasta las centésimas, el cálculo de porcentajes simples y la aplicación de la noción de proporcionalidad (regla de tres simple y directa). Además, establecen en forma adecuada las relaciones de equivalencia entre las principales unidades de longitud y tiempo expresadas con números decimales y fracciones; interpretan e identifican información estadística en diversas representaciones (gráfica, icónica y tabular); y calculan el perímetro y área de figuras planas que implican la composición y descomposición de figuras geométricas. Asimismo, identifican y diferencian, a partir de sus propiedades, figuras geométricas planas y del espacio.	A partir del conjunto de preguntas que estos estudiantes responden correctamente, se puede afirmar que resuelven situaciones problemáticas sencillas, rutinarias y no rutinarias, en las que el enunciado contiene, de manera explícita, la información necesaria y suficiente. La estrategia de solución consiste en la aplicación de una secuencia de dos operaciones aritméticas básicas o de algoritmos empleados en otros contenidos matemáticos (como la noción de proporcionalidad entre magnitudes, o la equivalencia de medidas de longitud y tiempo expresadas solo en números naturales). Los estudiantes en este nivel se están iniciando en la comprensión de los números racionales; manejan la noción de fracción como parte de un todo. Además, establecen la equivalencia entre denominaciones de nuestro sistema monetario que requieren el empleo de algunos números decimales. Asimismo, resuelven operaciones combinadas, con un nivel de signos de agrupación, con números naturales que implican respetar la jerarquía de las operaciones. También calculan el resultado de operaciones combinadas de adición y sustracción con fracciones homogéneas y con un nivel de signos de agrupación. Respecto del manejo de la geometría, identifican formas geométricas elementales a partir de sus propiedades (número de lados, congruencia de lados, medidas de sus ángulos). En cuanto a la estadística, interpretan y representan información en diagramas de barras a escala y en cuadros de doble entrada y hallan la frecuencia absoluta, determinando el total de un conjunto de datos.	Los estudiantes ubicados en el nivel previo resuelven problemas rutinarios que representan situaciones cotidianas con enunciados en los que la información necesaria para su resolución es explícita. La estrategia de solución para resolver estos problemas demanda aplicar una operación aritmética básica (adición, sustracción, multiplicación o división) o un procedimiento relacionado con la lectura, escritura y comparación de números naturales, con el establecimiento de relaciones de equivalencia entre las principales unidades de medida de tiempo, o con la interpretación y elaboración de cuadros y diagramas de barras con representación icónica. Asimismo, tienen un manejo solvente de los algoritmos convencionales de las cuatro operaciones básicas en el conjunto de los números naturales. Representan de diferentes maneras fracciones elementales y porcentajes de uso más frecuente (0%, 50% y 100%) y establecen equivalencias entre las diferentes unidades de longitud y tiempo. En cuanto al manejo de la geometría, identifican la forma de las figuras geométricas como un todo, pues solo identifican objetos a partir de su apariencia física (triángulo, cuadrado, rectángulo, rombo, cubo, cilindro, pirámide) y de su semejanza con objetos concretos. Respecto de la gestión de la información, leen y extraen datos en los casos más elementales (identificación de frecuencia absoluta en diagramas de barras y cuadros de doble entrada a escala de 1:1 con un número reducido de datos) dentro de un contexto personal y familiar, en el que el enunciado presenta en forma explícita la información necesaria y suficiente para su solución.

Cuadro 2.13. Ejemplos de preguntas de Matemática: Sexto grado de primaria

Ejemplos de preguntas																	
Se les presenta a los estudiantes una situación cotidiana referida a un fenómeno climático donde se indican los días y horas de lluvia de un determinado lugar. Puede observarse que la información no es solo textual, sino que está apoyada en información representada en un cuadro de doble entrada que indica la frecuencia de tiempo de lluvia en la región de Madre de Dios. El número de horas está representado por el dibujo de nubes, cada una de las cuales representa dos horas de lluvia, tal como lo indica la leyenda. El estudiante debe representar las horas de lluvia los días sábado y domingo, para lo cual tiene que interpretar la representación icónica, para luego poder completar el cuadro con la información proporcionada de forma textual.	<table><tr><th>Día</th><th>Número de horas</th></tr><tr><td>Lunes</td><td>  </td></tr><tr><td>Martes</td><td>  </td></tr><tr><td>Miércoles</td><td>  </td></tr><tr><td>Jueves</td><td> </td></tr><tr><td>Viernes</td><td></td></tr><tr><td>Sábado</td><td></td></tr><tr><td>Domingo</td><td></td></tr></table> Felipe fue a comprar con S/. 5. Lo que compró fue: 1 chupete a S/. 0,50 1 chocolate a S/. 1,20 1 chicle a S/. 0,30 1 gaseosa a S/. 1,30 ¿Cuánto le dieron de vuelto? Resuelve: $0,95 + 2,7 =$	Día	Número de horas	Lunes	  	Martes	  	Miércoles	  	Jueves	 	Viernes		Sábado		Domingo	
Día	Número de horas																
Lunes	  																
Martes	  																
Miércoles	  																
Jueves	 																
Viernes																	
Sábado																	
Domingo																	
Se les presenta a los estudiantes una situación problemática cotidiana que consiste en la compra de cuatro artículos diferentes, para que finalmente calculen el vuelto que recibirán. Este problema demanda que los estudiantes: interpreten y comprendan el enunciado, identifiquen las condiciones del problema y lo solicitado para relacionarlos entre sí y buscar una estrategia de solución adecuada. Se requiere hacer uso de las operaciones aritméticas básicas con números decimales. En el enunciado del problema se presentan explícitamente todos los datos necesarios y suficientes para responder la pregunta.	Se presenta a los estudiantes una situación intramatemática que involucra la operación de adición en números decimales con distinta cantidad de cifras decimales. Para resolverla, es necesaria la comprensión del significado de número decimal y del valor posicional de los números decimales, en particular, décimas y centésimas, para poder sumar correctamente teniendo en cuenta este valor posicional.																

Niveles	Puntajes
Suficiente	Entre 368 y 277
Básico	Entre 276 y 257
Previo	256 o menos

347 →

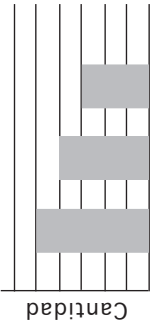
308 →

280 →

Cuadro 2.14. Niveles de desempeño en Matemática: Tercer grado de secundaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en este nivel resuelven situaciones problemáticas rutinarias y no rutinarias que demandan elaborar una secuencia de hasta tres operaciones aritméticas o utilizar la proporcionalidad directa con números racionales en su representación decimal. Estos estudiantes manejan adecuadamente las operaciones combinadas con números enteros respetando el orden jerárquico de dichas operaciones y se están iniciando en la comprensión de los números racionales. Asimismo, calculan el valor de la incógnita en ecuaciones e inecuaciones de primer grado de hasta tres pasos, con un nivel de signos de agrupación, definidas en el conjunto de los números racionales en su representación decimal. Calculan el conjunto solución de sistemas de dos ecuaciones con dos incógnitas definidos en el conjunto de los números enteros, cuando son presentados de manera ordenada. Sin embargo, solo resuelven situaciones problemáticas que demandan plantear una ecuación elemental de primer grado en los enteros. Se han iniciado en el desarrollo de la noción de función como una regla o una fórmula, lo cual les permite calcular valores puntuales de las imágenes de funciones polinómicas de primer y segundo grados, e identificar la expresión analítica que corresponde a un gráfico dado. También, resuelven situaciones problemáticas sencillas que demandan calcular el área del rectángulo y la noción de proporcionalidad geométrica. Elaboran diagramas de barras y resuelven situaciones problemáticas que requieren interpretar y deducir información presentada mediante tablas y diagramas estadísticos.	Los estudiantes que se encuentran en este nivel resuelven situaciones problemáticas rutinarias que demandan reproducir una secuencia de hasta tres operaciones aritméticas básicas con números naturales o con decimales positivos que tienen el mismo número de cifras decimales. Evidencian un adecuado manejo de los números naturales y emplean a consolidar su comprensión de los números enteros. Calculan el valor de la incógnita de ecuaciones de primer grado de menos de cinco términos, sin signos de agrupación, que demandan dos o tres pasos y están definidas en el conjunto de los números enteros. Interpretan las variables como una incógnita, es decir, como «la letra» que representa una cantidad desconocida que se debe hallar, lo cual no les permite aún manejar la noción de función. Además, los estudiantes ubicados en este nivel resuelven situaciones problemáticas sencillas que demandan la descomposición o integración visual de figuras geométricas simples. Grafican figuras geométricas elementales a partir de características y propiedades dadas. Emplean la simbología geométrica elemental. Interpretan diagramas de barras y cuadros de frecuencia de doble entrada. Resuelven situaciones problemáticas elementales que demandan identificar la frecuencia acumulada de un conjunto de datos.	Los estudiantes pertenecientes a este nivel resuelven situaciones problemáticas rutinarias que demandan reproducir algoritmos que consisten en una operación aritmética en el conjunto de los números naturales. Asimismo, se están iniciando en la comprensión de los números enteros, pues los representan gráficamente y calculan con ellos operaciones aritméticas básicas y operaciones combinadas que no demandan el reconocimiento de la jerarquía de las operaciones. Recodifican situaciones presentadas en lenguaje verbal utilizando una ecuación de primer grado en los casos más elementales (cuando el enunciado permite ir escribiendo la ecuación mediante la «traducción palabra por palabra», según se va leyendo). Identifican las figuras geométricas básicas y sus elementos a partir de su representación gráfica. Interpretan solo los diagramas de barras más elementales y comparan frecuencias absolutas.

Cuadro 2.15. Ejemplos de preguntas de Matemática: Tercer grado de secundaria

Ejemplos de preguntas	
Se propone una situación problemática de enunciado verbal no rutinaria en la que se presenta un contexto fantástico: mago que aumenta la cantidad de dinero que tienen las personas, cobrando por ello, y se da la cantidad final que tiene uno de sus clientes. Se le pide al estudiante que halle la cantidad de dinero que tenía inicialmente dicho cliente. Para responder a esta pregunta, el estudiante debe, en primer lugar, comprender la situación y elaborar una estrategia de solución que le permita relacionar los datos y condiciones del problema: el dinero que tiene al final el cliente y las distintas transformaciones que se realizaron. Por ejemplo, el estudiante podría elegir una secuencia de operaciones aritméticas que reproduzcan las transformaciones pedidas, o, utilizar una estrategia algebraica, tal como plantear una ecuación lineal. Seguidamente deberá desarrollar la estrategia que ha elegido para calcular el monto del dinero pedido y, finalmente, interpretar el resultado según lo que le solicita la pregunta. Se propone una situación problemática no rutinaria sobre el registro del número de gallinas. La información se complementa con un diagrama de barras que presenta el número de gallinas según su color. Se da solo el número de gallinas de una de las clases y se le pide al estudiante que calcule el número total de las gallinas de las otras clases. Esta pregunta evalúa la capacidad de resolución de problemas, en particular, la interpretación del diagrama de barras y el empleo de la proporcionalidad geométrica. Para responder a la pregunta, el estudiante debe, en primer lugar, interpretar la situación —y en particular el diagrama de barras—, para lo que deberá identificar la barra de la cual se da el dato numérico. Seguidamente, deberá definir un paso intermedio no explícito que le permita hallar la medida de la escala gráfica utilizada en el diagrama de barras y utilizar alguna estrategia que le permita calcularla. Finalmente, deberá utilizar ese dato para calcular el número total de gallinas que se le pide. Se propone una situación problemática no rutinaria sobre dos opciones para la compra de revistas y lo que ocurre con el dinero que se tiene a la luz de cada opción. Se presenta también la ecuación lineal que representa dicha situación. Se le pide al estudiante que interprete el significado de la incógnita de la ecuación presentada. Esta pregunta evalúa la capacidad de comunicación matemática, en particular, el manejo de la noción de variable algebraica entendida como incógnita. Para responderla, el estudiante deberá comprender la situación, integrando la información —tanto la parte presentada mediante lenguaje verbal como la presentada utilizando el lenguaje algebraico. Seguidamente deberá establecer las relaciones entre ambas representaciones para atribuir el significado que se le otorga a la incógnita dentro del contexto de la pregunta.	Un mago “milagroso” duplica el dinero, con la condición de que por cada “milagro” que realice, se le entregue 40 nuevos soles. Si una persona después de pagar un “milagro” tiene 100 nuevos soles, ¿cuánto dinero tenía inicialmente? En la casa de Pablo crían gallinas blancas (B), negras (N) y rojas (R). Pablo hace el diagrama que se muestra abajo, para representar la cantidad de gallinas de cada color.  A partir del diagrama, responde: Si hay 15 gallinas blancas, ¿cuántas n o son blancas? a) 7 b) menos de 15 c) 21 d) 30
	Arturo quiere comprar una colección de revistas de música (cada revista tiene el mismo precio). Si decide comprar doce revistas con el dinero que tiene, le faltarian S/. 6 pero si compra ocho revistas, entonces le sobran S/. 6. La siguiente ecuación representa la situación propuesta: $12x - 6 = 8x + 6$ En la ecuación anterior, la incógnita “x” representa: a) El costo total de las revistas que compró. b) La cantidad de dinero que tiene Arturo. c) La cantidad de revistas que compró. d) El costo de cada revista.

598

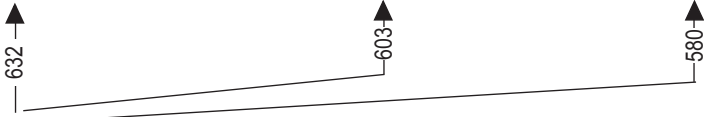
558

522

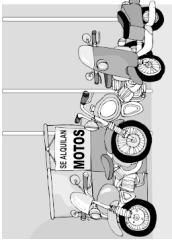
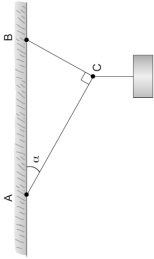
Cuadro 2.16. Niveles de desempeño en Matemática: Quinto grado de secundaria

Nivel Suficiente	Nivel Básico	Nivel Previo
Los estudiantes ubicados en el nivel suficiente pueden enfrentar situaciones matemáticas novedosas de diversos tipos, realizar estimaciones adecuadas de longitud y tiempo, interpretar información presentada en diversos diagramas, discriminar la pertinencia de la información que se les presenta y hacer inferencias sencillas que les permiten tomar decisiones adecuadas. Mediante la adaptación o la construcción de estrategias, estos estudiantes resuelven situaciones problemáticas rutinarias y no rutinarias, en diversos contextos, utilizando notaciones y términos matemáticos convencionales. Los hechos y conceptos son elementos de redes más amplias que les permiten establecer conexiones en distintos niveles y contextos. Estos estudiantes justifican y explican sus procedimientos y razonamientos mediante un lenguaje matemático formal acorde con el grado. Los estudiantes que se ubican en este nivel manejan adecuadamente el conjunto de los números racionales utilizando cualquiera de sus representaciones. Han logrado consolidar su razonamiento proporcional, pues resuelven situaciones problemáticas utilizando nociones de proporcionalidad con dos o más magnitudes. Resuelven situaciones problemáticas que demandan plantear ecuaciones algebraicas lineales con una incógnita, de tres o cuatro términos, con un nivel de signos de agrupación y con coeficientes en los racionales. Utilizan las variables para representar números, patrones y dependencias funcionales (es decir, para representar argumentos en funciones). Esto les permite manejar la noción de función como una correspondencia entre dos variables. Asimismo, empiezan a comprender el carácter predictivo de las funciones. Respecto del pensamiento geométrico, estos estudiantes manejan los procedimientos y los conceptos básicos de la geometría analítica, identifican y grafican ecuaciones de rectas. Resuelven problemas que demandan el empleo de triángulos notables y de razones trigonométricas (seno, coseno y tangente). Además, resuelven situaciones problemáticas que involucran interpretar información presentada en los diversos formatos usuales de la estadística. También resuelven situaciones problemáticas que involucran el cálculo y la comparación de probabilidades de eventos sencillos.	Los estudiantes ubicados en el nivel básico, mediante la adaptación y reproducción de estrategias previamente aprendidas, resuelven problemas rutinarios y no rutinarios que demandan interpretar situaciones en diversos contextos y que, a su vez, emplean en el enunciado algunas notaciones y términos matemáticos convencionales. Asimismo, resuelven situaciones comerciales que involucran cálculos de descuentos e interés simple, realizan estimaciones de medidas de longitudes pequeñas, siguen y ejecutan instrucciones verbales breves y analizan con sentido crítico información presentada en diversos diagramas. Además, estos estudiantes realizan cálculos aritméticos con operaciones básicas en el conjunto de los números racionales (representación fraccionaria y decimal) y en el conjunto de los números enteros. Asimismo, resuelven problemas de texto en situaciones comerciales que involucran el significado de los números negativos. Estos estudiantes utilizan tanto la noción de incógnita como la de variable. Esto les permite manejar en un nivel intuitivo la noción de función, entendida como regla de correspondencia entre dos magnitudes relacionadas. Además, resuelven situaciones realistas que demandan plantear ecuaciones lineales enteras con una incógnita, hasta con tres términos y sin signos de agrupación. Su pensamiento proporcional se encuentra en proceso de construcción pues solo resuelven problemas utilizando la proporcionalidad simple y el cálculo directo de porcentajes. Identifican elementos y propiedades de figuras geométricas básicas y utilizan el Teorema de Pitágoras para sustentar afirmaciones. Calculan áreas de figuras geométricas básicas y de prismas rectos. Los estudiantes de este nivel manejan las definiciones de las razones trigonométricas en el triángulo rectángulo de manera incipiente e inconexa. Asimismo, interpretan y grafican información estadística presentada en tablas de doble entrada y en diagramas de barras rectangulares, calculan medidas de tendencia central en un listado ordenado de datos y la probabilidad de ocurrencia de eventos sencillos.	Los estudiantes que se encuentran en el nivel previo resuelven situaciones problemáticas rutinarias y sencillas de tipo comercial que requieren solo la aplicación de las cuatro operaciones básicas —y combinaciones de estas— y que no requieren el uso del porcentaje. También pueden realizar algunas estimaciones de medida de longitudes pequeñas, seguir y ejecutar instrucciones verbales breves y directas, y leer información presentada en tablas o en diagramas sencillos, sin evidenciar la capacidad crítica para analizar la pertinencia de la información que reciben. Se ha observado que únicamente resuelven situaciones en las que toda la información relevante aparece explícita en el enunciado y que solo demandan reproducir procedimientos breves previamente aprendidos. Respecto del manejo de información cuantitativa, comparan información a partir de la lectura directa de frecuencias presentadas en tablas y diagramas de barras. Calculan una secuencia de hasta tres operaciones aritméticas básicas en el conjunto de los números naturales. Su pensamiento proporcional es incipiente. Además, representan situaciones de enunciado verbal mediante el uso del lenguaje algebraico en los casos más elementales (cuando el enunciado permite ir escribiendo la ecuación por medio de la «traducción palabra por palabra»). También calculan el valor de la incógnita en ecuaciones lineales enteras con tres términos. Utilizan las letras únicamente como incógnitas, lo que no les permite aún manejar la noción de función —que exige el uso de las letras como variables para representar argumentos. Identifican los elementos y algunas propiedades de figuras geométricas básicas aunque no relacionan estas propiedades entre sí. Asimismo, manejan una simbología geométrica elemental.

Niveles	Puntajes
Suficiente	Entre 635 y 554
Básico	Entre 553 y 501
Previo	500 o menos



Cuadro 2.17. Ejemplos de preguntas de Matemática: Quinto grado de secundaria

Ejemplos de preguntas	
Esta situación problemática demanda que los estudiantes modelen la situación asignando un objeto matemático que la represente. Para esto deben interpretar el enunciado y relacionar las variables (costo de alquiler por un día y la distancia total recorrida por la moto), para finalmente representarla de manera generalizada para un recorrido de «x» kilómetros, haciendo uso del lenguaje algebraico. El estudiante debe tener la noción de variable de tal forma que esta le permita representar las dependencias y relaciones pedidas en la situación.	En una ciudad, el alquiler de motos cuesta S/. 20 el día más S/. 0,50 por kilómetro recorrido. Halla la expresión que representa el costo de alquilar una moto durante un día para un recorrido de x kilómetros. 
Se les presentó a los estudiantes una situación que presenta un enunciado verbal apoyado en un soporte gráfico. Esta situación involucra el empleo de la razón trigonométrica «seno» para usar proporciones o establecer relaciones entre las medidas de los lados de triángulos rectángulos semejantes. Para resolver este problema el estudiante debe interpretar la situación presentada en el enunciado y relacionarla con el soporte gráfico dado y debe identificar los elementos del triángulo rectángulo de tal forma que pueda establecer una proporción para relacionar las longitudes de los lados involucrados con la razón trigonométrica dada.	Se sostiene un bloque del techo mediante cuerdas como indica la figura. Si $\text{sen } a = \frac{1}{4}$ y la cuerda CB mide 30 cm, calcula la distancia entre A y B. 
Se les presentó a los estudiantes una situación intramatemática, en la que se muestra un conjunto de números racionales (representación decimal) para graficar en la recta numérica y seleccionar el más próximo al origen. Esta pregunta demanda a los estudiantes comparar y establecer relaciones de orden entre los números dados y establecer cual es el más próximo al origen.	Ubica los siguientes números en una recta numérica y señala el más próximo al origen. -0,36; 0,48; -2; 1 y -0,53

Después de haber señalado las principales características de la EN 2004, en el siguiente capítulo se dará cuenta del marco conceptual en el cual se inscriben las pruebas de rendimiento de las áreas evaluadas. Se describirá las dimensiones del enfoque de evaluación que se ha asumido tanto para las preguntas de Comprensión de textos escritos como para las de Matemática.



Nuestro sistema educativo propone en el currículo capacidades fundamentales que orientan el desarrollo de las áreas curriculares para la población estudiantil del país. El desarrollo de dichas capacidades proporciona a los estudiantes importantes herramientas que les servirán para propiciar un aprendizaje continuo y autónomo y que son esenciales para el pleno ejercicio de su ciudadanía. Estas capacidades —tales como analizar, inferir, razonar, saber expresarse, argumentar y resolver problemas— son evaluadas en la EN 2004 desde las áreas de Comunicación y Matemática.

La EN 2004 recoge información acerca del nivel de desarrollo de estas capacidades en los estudiantes por medio de pruebas de rendimiento³⁹ que proponen a los estudiantes diversas tareas que son similares a las que encontrarán tanto en contextos reales como en situaciones propias del área.

En tal sentido, la EN 2004, en comparación con las evaluaciones anteriores, constituye una aproximación más cercana a la evaluación de estas capacidades en tanto el enfoque de área evalúa no solamente el grado de aprendizaje de los contenidos curriculares, sino también, y principalmente, evalúa en qué medida los estudiantes cuentan con las herramientas básicas para desempeñarse adecuadamente en la escuela y en otras esferas de su vida cotidiana.

En efecto, para poder afirmar que un estudiante muestra un nivel de desarrollo de las capacidades apropiado para cierta área, se requiere que las pruebas recojan información compleja y variada.⁴⁰ Para dar cuenta de estos niveles de logro, es necesario plantear una diversidad de preguntas que impliquen en su resolución la puesta en uso de diversas capacidades, el manejo de ciertos contenidos y el desenvolvimiento en diferentes contextos. Por esta razón, el enfoque de evaluación de cada área de la EN 2004 se organiza en tres dimensiones:

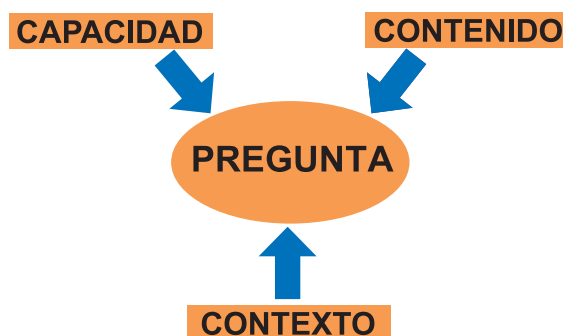
- *Las capacidades:* El estudiante requiere poner en práctica, al enfrentarse a las situaciones que se le presentan dentro y fuera de la escuela, habilidades tales como representar, razonar, analizar o resolver problemas. En el caso particular de las pruebas, las capacidades evaluadas son aquellas que el estudiante requiere para enfrentar exitosamente a las preguntas propuestas.

38. Para más detalle sobre el contenido de este capítulo puede consultarse el documento Marco de Trabajo de las Pruebas de Rendimiento de la EN 2004 disponible en: <http://www.minedu.gob.pe/umc/2004/marc-trab/MarcTrabPruebEN2004.pdf>.

39. Para elaborar las pruebas se han seleccionado las capacidades más importantes y susceptibles de ser evaluadas mediante pruebas estandarizadas de papel y lápiz.

40. Ello ha sido posible no solo por la aproximación funcional que se ha asumido como marco evaluativo, sino también porque se ha incluido en las pruebas un mayor número de preguntas en las que los estudiantes deben construir sus propias respuestas, lo que ha permitido evaluar los procesos que siguen y las estrategias que utilizan.

- *Los contenidos:* Están referidos a los conocimientos específicos que el estudiante debe manejar en cada una de las áreas evaluadas. Dichos contenidos han sido tomados de los currículos vigentes al momento de diseñar las pruebas.
- *Los contextos:* Son aquellos escenarios en los que las capacidades y contenidos deben ser puestos en práctica. Estos contextos pueden ser propios de la vida escolar, personal o pública.



Cada una de las preguntas de la prueba plantea al estudiante una tarea en la que se interrelaciona una capacidad —requerida por el estudiante para resolverla— y un contenido —que el estudiante debe conocer—, en un contexto determinado. La graduación de cada una de las tres dimensiones de este modelo ha permitido generar una diversidad de tareas que abarca los diferentes niveles de dominio de los estudiantes en el área evaluada.

La propuesta de evaluación de la EN 2004 busca estimar la habilidad de los estudiantes al enfrentarse a tareas de distinta complejidad y construir una escala que permite identificar y describir lo que los estudiantes pueden hacer, desde los niveles de dominio más sencillos hasta los más complejos. Dicha escala ha hecho posible reportar los diferentes niveles de dominio que muestran los estudiantes con respecto a lo que se espera de ellos en cada uno de los grados evaluados.

A continuación, se presentan las dimensiones del modelo de evaluación de la EN 2004 desde la perspectiva particular de cada una de las áreas evaluadas.

3.1. Dimensiones del modelo de evaluación en Comprensión de textos escritos⁴¹

La comprensión de textos es entendida como un proceso en el que el lector utiliza sus conocimientos previos (conocimientos del mundo, del lenguaje, pautas culturales) para interactuar con el texto y reconstruir su sentido.

41. Como ya se señaló en el capítulo anterior, en el área de Comunicación se evaluó tanto la comprensión como la producción de textos escritos. En el presente informe solo se presentarán los resultados sobre el desempeño de los estudiantes en la prueba de Comprensión de textos escritos. La producción escrita será tema de otro reporte.

La lectura, entonces, no se reduce a un simple proceso de decodificación de palabras, frases y oraciones; de identificación de los significados de palabras; y de reconocimiento de estructuras gramaticales: la lectura depende también del desenvolvimiento de un conjunto de otras habilidades. El lector extrae la información para deducir las ideas implícitas, y se plantea preguntas que le permiten reflexionar sobre los significados que va construyendo.

La EN 2004 precisamente busca explorar los procesos que el estudiante realiza al interpretar un texto en una situación concreta de comunicación, en este caso, a partir de la lectura de diversos textos escritos.

El modelo de evaluación de la comprensión de textos considera las tres dimensiones señaladas anteriormente: capacidades, contenidos (textos) y contextos (uso de la lectura).

A. CAPACIDADES

Están referidas a las habilidades que el lector pondrá en juego al realizar las tareas que se le proponen y que pretenden simular los tipos de tareas que los estudiantes encontrarán en situaciones auténticas de lectura en la vida.

Para efectos de esta evaluación a escala de sistema, la prueba de Comprensión de textos escritos en el área de Comunicación está organizada alrededor de tres capacidades: *Obtiene información explícita*, *Hace inferencias*, y *Reflexiona y evalúa*. A continuación se explica brevemente cada una de estas capacidades.

- a) *Obtiene información explícita*: Esta capacidad permite al lector recuperar la información que se encuentra de manera literal en el texto. La recuperación efectiva requiere una comprensión relativamente inmediata o automática del texto, puesto que se necesita hacer poca o ninguna inferencia o interpretación.
- b) *Hace inferencias*: Esta capacidad permite al lector ir más allá de la superficie y deducir aquella información que es implícita y que es necesaria para comprender el texto en aspectos particulares y en su totalidad. En este caso, los lectores deben, principalmente, interpretar las relaciones entre las ideas explícitas e implícitas del texto tomando en cuenta el contexto comunicativo del mensaje, su propósito, su interlocutor.
- c) *Reflexiona y evalúa*: Por medio de esta capacidad, los lectores critican las ideas propuestas en el texto o los recursos utilizados por el autor para transmitir ese significado, y juzgan si son adecuados o no. El lector inmerso en este proceso se distancia del texto, lo considera objetivamente y evalúa su calidad y adecuación a partir de su contenido o su forma, desde una perspectiva crítica.

Para aproximarnos a las capacidades, la evaluación considera algunos desempeños relevantes asociados a cada capacidad. Un desempeño es la actuación del estudiante ante un estímulo determinado y permite inferir lo que este sabe y puede hacer de manera más precisa.⁴²

42. Para más información sobre los desempeños evaluados en el área de Comunicación, véase el Marco de Trabajo de las Pruebas de Rendimiento. En: <http://www.minedu.gob.pe/umc/2004/marctrab/MarcTrabPruebEN2004.pdf>.

B. CONTENIDOS (TEXTOS)

El texto es el entramado de ideas relacionadas que forman una unidad global de de sentido. Puede estar organizado en uno o más párrafos (cuentos, noticias, ensayos, etc.) o en una estructura gráfica notoriamente distinta de la prosa estándar (cuadros, avisos, afiches publicitarios, infografías, etc.). Desde el punto de vista de la función que cumplen, distinguiremos diversos tipos de texto propuestos en la evaluación:

- a) *Texto narrativo*: representa una secuencia de hechos que transcurren en el tiempo. Su propósito es crear mundos ficcionales en los que se relata hechos vividos por personajes reales o imaginarios en un periodo determinado. En la prueba, se presenta historietas breves, fábulas, leyendas, cuentos, etc.
- b) *Texto descriptivo*: ofrece información concreta acerca de una persona, animal, objeto, experiencia, o proceso. En este tipo de texto se presentan las características y circunstancias ordenadas en el espacio, de manera sincrónica. En la prueba, se utilizan textos descriptivos como recetas, instrucciones, cartas, crónicas, cuadros estadísticos, certificados, infografías, etc.
- c) *Texto expositivo o explicativo*: expone un tema o materia, la explica, informa acerca de algo. En la prueba se presentan artículos científicos, académicos y de otra índole social o cultural.
- d) *Texto argumentativo*: apoya o discrepa de una afirmación cuya validez es cuestionable o discutible. Todo texto argumentativo tiene como meta persuadir o convencer a la audiencia a la que se dirige acerca del valor de la tesis para la cual busca aprobación. En la prueba, se presenta cartas, textos de opinión, afiches publicitarios, ensayos.

Cabe resaltar que los tipos de texto anteriores pueden contener tanto información verbal como ícono-verbal.

C. CONTEXTOS (USOS DE LA LECTURA)

Se refiere a la finalidad para la que fue construido el texto. Podemos distinguir tres usos de lectura:

- a) *Recreativo*: El estudiante utiliza esta lectura para recrear mundos ficcionales. Generalmente, es un tipo de lectura privada.
- b) *Público*: El estudiante usa esta lectura para conocer las actividades políticas, sociales, culturales o económicas de la sociedad. Este tipo de lectura incluye el uso de documentos oficiales y de información acerca de acontecimientos públicos, como las noticias, comunicados oficiales, certificados, recibos, etc.
- c) *Educacional*: El estudiante utiliza esta lectura con un propósito instructivo relacionado con la adquisición de información como parte de una tarea de aprendizaje más amplia.

Cuadro 3.1: Dimensiones del modelo de evaluación en Comprensión de textos escritos

Capacidades	Contenidos (textos)	Contextos (uso de la lectura)
a. Obtiene información explícita	a. Texto narrativo	a. Recreativo
b. Hace inferencias	b. Texto descriptivo	b. Educacional
c. Reflexiona y evalúa	c. Texto expositivo o explicativo	c. Público
	d. Texto argumentativo	

3.2. Dimensiones del modelo de evaluación en el área de Matemática

Las pruebas del área de Matemática pretenden evaluar el dominio de habilidades y conocimientos matemáticos útiles para desempeñarse eficazmente ante situaciones problemáticas novedosas o rutinarias cuya solución requiera la puesta en práctica de dichas habilidades y conocimientos. Para poder evaluar este dominio, cada una de las tareas presentadas ha considerado las siguientes capacidades, contenidos y contextos.

A. CAPACIDADES

Las capacidades están referidas a las habilidades matemáticas complejas que el estudiante debe poner en práctica al enfrentarse a las preguntas de la prueba. Cada una de estas preguntas ha sido elaborada tratando de presentar situaciones que se le podrían presentar en su vida cotidiana, incluyendo la vida escolar.

En las pruebas de Matemática, se han considerado tres capacidades que a continuación son detalladas.

- a) *Resolución de problemas*: Para la EN 2004 se entiende como *situación problemática* a aquella que plantea una cuestión de contenido matemático inscrita en la matemática escolar y cuyo método de solución, por lo general, no es inmediatamente accesible al sujeto que intenta responderla. Por lo tanto este debe buscar y establecer relaciones para hacer frente a la situación que se le propone. En esta evaluación las situaciones problemáticas, presentadas todas ellas en formato impreso, están referidas a contextos matemáticos y a contextos propios de la vida real. Para resolverlas, el estudiante deberá comprender e interpretar la situación presentada, diseñar una estrategia novedosa o adaptar una ya conocida para resolverla, llevar a cabo su estrategia y, finalmente, interpretar sus resultados matemáticos dentro del contexto de la situación presentada.

Actualmente, la resolución de problemas es considerada como la razón de ser de la actividad matemática, principalmente por ser un medio eficiente para desarrollar el conocimiento matemático y un logro indispensable de una buena formación matemática.

Además, la elaboración de estrategias para resolver problemas genera confianza en los estudiantes al permitirles comprobar sus posibilidades de hacer matemática, estimula su autonomía y creatividad, expresa el grado de comprensión de los conocimientos y les facilita mecanismos de transferencia de lo aprendido a otras situaciones.

- b) *Comunicación matemática*: Se refiere a la capacidad del estudiante para interpretar, identificar, relacionar, clasificar, representar y recodificar tanto la información matemática, como la que él necesita producir para responder a distintas situaciones del medio. Es importante mencionar que la comunicación matemática va más allá del conocimiento de terminología o simbología especializada.

La matemática es un lenguaje poderoso y un medio para comprender y comunicar ideas de la ciencia y la tecnología. Permite presentar dichas ideas de forma verbal, gráfica o simbólica (o mediante una combinación de estas). Además, dicho lenguaje es una herramienta que promueve el desarrollo de las habilidades para interpretar y representar ideas matemáticas, y, adicionalmente, es el soporte tanto para elaborar justificaciones y argumentos plausibles y fundamentados como para integrar y relacionar información proveniente de diferentes fuentes (textos, diagramas, gráficos, etc.).

Por consiguiente, en esta capacidad, la EN 2004 pretende evaluar en qué medida el estudiante recibe y comunica información en forma clara y eficiente mediante el uso del lenguaje matemático y argumenta de manera fundamentada, sobre la base de conocimientos matemáticos, las afirmaciones que realiza.

- c) *Aplicación de algoritmos*: los algoritmos son un conjunto de acciones (operaciones y procedimientos), pasos secuenciales previamente establecidos y formas de actuar para llegar a resolver situaciones matemáticas. Se trata siempre de formas de proceder prefijadas, efectivas y sistemáticas, que se orientan al logro de un objetivo específico. Es muy importante poseer una fluidez operacional, es decir, tener y usar métodos eficientes y precisos para realizar algoritmos.

Es necesario evaluar en qué medida los estudiantes han aprendido a utilizar y seleccionar determinados algoritmos para poder enfrentarse adecuadamente a las situaciones que se le presentan. En la EN 2004 se ha incluido la aplicación de algoritmos, porque constituye un conjunto de herramientas útil para solucionar una diversidad de problemas. Las preguntas que forman parte de la EN 2004 que evalúan esta capacidad intentan reproducir situaciones a las que el estudiante ya tuvo oportunidad de enfrentarse y pretenden evaluar en qué medida el estudiante utiliza y adapta algoritmos para resolver ejercicios matemáticos. Se evalúa la aplicación de algoritmos aritméticos, algebraicos, geométricos y estadísticos.

Para poder aproximarnos a las tres capacidades evaluadas en el área de Matemática, han sido considerados algunas capacidades específicas y relevantes asociadas a cada una de las tres capacidades descritas previamente. Estas capacidades específicas pueden ser entendidas como las habilidades necesarias para la adecuada actuación del estudiante ante una situación determinada y que permite inferir lo que sabe y puede hacer.⁴³

43. Dichas capacidades específicas, en el caso del nivel primario, han sido adaptadas en su mayoría de las que presenta la ECB —vigente en el momento de diseñar la evaluación— para el ciclo correspondiente al grado evaluado. En el caso de secundaria, han sido elaboradas a partir de las capacidades y contenidos contemplados en los DCB vigentes al momento de elaborar la prueba. Para más información sobre los desempeños evaluados en el área de Matemática, véase el documento Marco de Trabajo de las Pruebas de Rendimiento en: <http://www.minedu.gob.pe/umc/2004/marctrab/MarcTrabPruebEN2004.pdf>.

B. CONTENIDOS

Se refieren a los conocimientos matemáticos que se evalúan en la EN 2004 y que varían de acuerdo con el grado evaluado. Estos contenidos han sido seleccionados de los currículos vigentes en el momento de la elaboración de las pruebas. Dichos contenidos han sido reagrupados tratando de considerar el enfoque fenomenológico⁴⁴ en cuatro categorías: Número y cantidad, Álgebra y funciones, Espacio y forma, y Estadística y probabilidad.

Seguidamente, se detalla cada una de estas categorías.

- a) *Número y cantidad*: Este contenido está muy ligado históricamente con la matemática escolar y se refiere a los diferentes conjuntos numéricos en todas sus representaciones —enteros, fracciones, decimales— y sus propiedades, como la divisibilidad, por ejemplo; las operaciones aritméticas, tanto elementales como combinadas; las propiedades de dichas operaciones; y la proporcionalidad y sus aplicaciones, como regla de tres y porcentaje.
- b) *Álgebra y funciones*: El álgebra puede ser definida como el arte de descubrir leyes o patrones que rigen la relación entre diversas variables, independientemente de los valores involucrados. Permite transformar expresiones matemáticas complejas en otras equivalentes y más sencillas. Es una herramienta poderosa, pues hace posible generalizar situaciones diversas mediante la simbolización. Ofrece, asimismo, un medio para trabajar con conceptos en un nivel abstracto que facilitará su comprensión y sus relaciones; además, posibilita la interiorización de estos conceptos, al ser un medio adecuado para representarlos.

En la EN 2004, este contenido se evalúa solo en el nivel secundario y se ha considerado para ello funciones lineales y cuadráticas, ecuaciones, inecuaciones, sistemas de ecuaciones e introducción a la geometría analítica.

- c) *Espacio y forma*: El manejo del espacio y de la forma provee a las personas de herramientas, tanto para realizar, representar, organizar y estructurar sus actividades y movimientos, como para organizar su pensamiento y su razonamiento. El logro de esta estructuración será posible si el sujeto es capaz de comprender las propiedades, posiciones, representaciones y transformaciones de los objetos y las relaciones entre los mismos, lo cual implica comprender las relaciones entre las formas y representaciones visuales.

En la EN 2004, se consideran los aspectos espaciales de los sólidos y de las figuras (objetos geométricos); sus elementos, relaciones y propiedades; la representación e interpretación de objetos tridimensionales en el plano; los patrones geométricos como modelos de muchos objetos y fenómenos; el manejo del vocabulario geométrico básico estandarizado; la orientación espacial; la comparación de figuras geométricas; el cálculo de medidas tales como el área, volumen, perímetro; y las razones trigonométricas.

- d) *Estadística y probabilidad*: La estadística permite organizar, representar y realizar el análisis e interpretación de datos para elaborar conclusiones y tomar decisiones sobre una base científica. También posibilita la predicción de los comportamientos de determinados fenómenos.

44. En el enfoque fenomenológico propuesto por Steen (1998), se propone agrupar los contenidos a partir de cómo se presentan naturalmente en los problemas de la realidad.

En la EN 2004, se evalúa la recolección, organización, presentación y análisis de datos. La probabilidad es objeto de evaluación solo en quinto de secundaria, mediante preguntas que abordan la cuantificación y representación de la variabilidad, la posibilidad y la probabilidad de ocurrencia de eventos.

C. CONTEXTOS

Un aspecto importante de la formación matemática es el desarrollo en el estudiante de las capacidades que le permitan usar y hacer matemática en una gran variedad de situaciones. Por ello, la comprensión y la intuición matemática de los estudiantes necesitan ser evaluadas en diversos escenarios de acción.

La EN 2004 ha considerado dos tipos de contexto: el intramatemático y el extramatemático.

- a) *Intramatemático*: Trata específicamente las tareas del universo matemático, utiliza sus propios símbolos y presentan objetos matemáticos sin referirlos o inscribirlos en el mundo real.
- b) *Extramatemático*: Ubica los objetos matemáticos en relación con los diversos aspectos de la vida del estudiante. Pueden clasificarse, a su vez, de acuerdo con el nivel de uso o de familiaridad que tengan para este. Lo más cercano a él suele ser más sencillo y susceptible de ser representado mentalmente por el estudiante. Consideramos así el contexto familiar como el más cercano, seguido del contexto escolar, laboral, local, nacional y, científico y tecnológico. De esta manera, es posible determinar una graduación más o menos continua de situaciones.

Independientemente de la distancia entre las situaciones y los estudiantes, se pretende que las tareas estén basadas en contextos auténticos, es decir, que probablemente ocurran en el mundo real. Si la educación matemática debe preparar a los estudiantes para que sean ciudadanos activos e informados, tiene que tratar con contextos reales tales como los que se presentan en las noticias, las situaciones de compra-venta y de presupuestos para realizar un trabajo. Esto no excluye, sin embargo, contextos ficticios o artificiales basados en representaciones estilizadas de problemas.

Cuadro 3.2: Dimensiones del modelo de evaluación en Matemática

Capacidades	Contenidos	Contextos
a. Resolución de problemas	a. Número y cantidad	a. Intramatemático
b. Comunicación matemática	b. Álgebra y funciones	b. Extramatemático
c. Aplicación de algoritmos	c. Espacio y forma	
	d. Estadística y probabilidad	

En el siguiente capítulo se describirán algunas características del sistema educativo de nuestro país a través de indicadores como la tasa de cobertura, de eficiencia, de conclusión, etc. correspondientes a los grados y niveles evaluados. Consideramos que estos aspectos permitirán contextualizar y dar sentido a los resultados de aprendizaje obtenidos por los estudiantes evaluados que se presentarán en el capítulo 5.



Antes de presentar los resultados de la EN 2004, resulta importante atender a algunos indicadores educativos con el fin de tener una mirada integral sobre el desempeño del sistema educativo peruano. En efecto, el sistema educativo puede ser visto desde diversas perspectivas que comprenden un conjunto bastante amplio y complejo de fenómenos, dimensiones y problemas.

Si partimos de considerar que la educación debería ser uno de los mecanismos más eficaces para evitar la reproducción de las desigualdades socioeconómicas, una de las tareas principales de la política educativa debería ser la distribución equitativa de las oportunidades para las personas en la sociedad. Por consiguiente, las políticas educativas deben contemplar un componente de equidad desde su propia definición. De allí que la equidad no sea accesoria o adicional al sistema sino que hace directamente a su calidad. Esta calidad, por su parte, exige que los servicios educativos sean relevantes para el desarrollo de las personas y su potencial, y que se tomen en cuenta sus especificidades y los desafíos de la sociedad contemporánea. La calidad también requiere que los sistemas educativos sean pertinentes para las circunstancias particulares de los estudiantes, de modo que la diversidad sea un factor de enriquecimiento en vez de un criterio de exclusión (Machinea, Bárcena y León 2005).

En este sentido, un sistema educativo debe ser capaz no solo de garantizar igualdad de oportunidades a las personas para aprender, sino también debe hacer que ese aprendizaje sea accesible para todos los miembros de una determinada sociedad, de manera que las personas puedan ingresar, permanecer, progresar y lograr los objetivos que se plantea este.

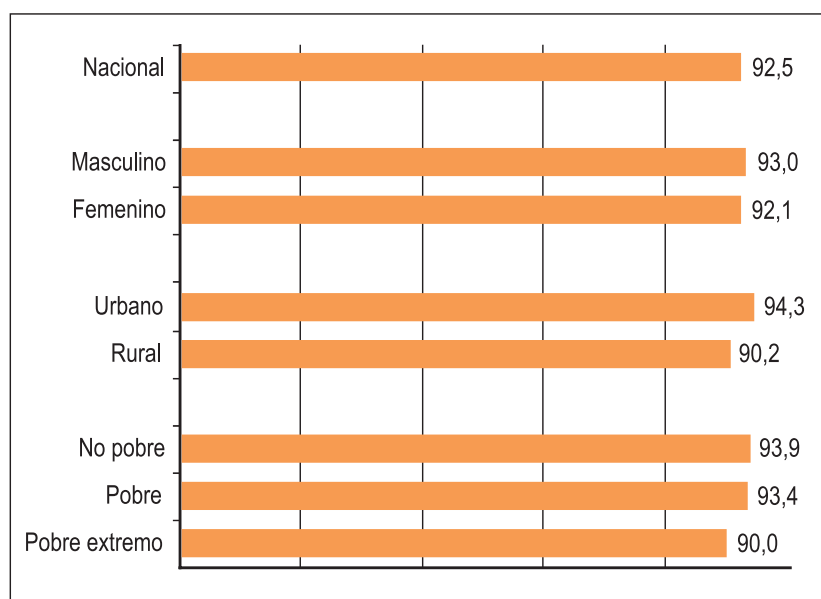
En el presente capítulo se intentará aproximarse, de manera breve, a través de algunos indicadores de acceso, flujo y conclusión, dar cuenta de la situación en que se inscriben los indicadores de rendimiento que se presentan en el siguiente capítulo. Para ello, se tomarán los principales indicadores que han sido consensuados en los principales planes de educación.

4.1. Educación Primaria: acceso, tránsito y conclusión

El servicio educativo primario constituye probablemente la organización social más extensa del país, con más de 34 mil instituciones educativas que, en su conjunto, atienden a casi todos los niños en edad escolar. Esta extensión responde al asentamiento poblacional disperso que caracteriza a nuestro país.⁴⁵

45. De acuerdo al Censo de Población y Vivienda de 1993, en el Perú existían 84 046 centros poblados, de los cuales 55 mil tenían menos de 100 habitantes (MED 2001).

Grafico 4.1: Tasa de cobertura en nivel de la población de 6 a 11 años, según sexo, área y nivel de pobreza



Como ha sido señalado por varios estudios, al interior de nuestro sistema educativo, el acceso de la población de 6 a 11 años a la escuela primaria ha alcanzado cifras cercanas a la universalización (93%), sin presentar grandes brechas por género, área de residencia o nivel de pobreza, como se puede observar en el gráfico 4.1.

Fuente: MED (2005b).

Este logro en el acceso a la educación primaria estaría asociado, entre otros factores, a la importante valoración social de la educación que a lo largo de todo el siglo XX, y con mayor énfasis a partir de la segunda mitad, ha mostrado nuestra población y, vinculado a este aspecto, como se ha señalado, la presencia de una extensa red de instituciones educativas (MED 2001). En efecto, la demanda de las zonas rurales más alejadas y de escasa población por la educación primaria se ha podido cubrir a través de las llamadas escuelas multigrado y unidocentes⁴⁶ gestionadas en la mayor parte de las veces por los propios pobladores de dichas zonas. No obstante, el logro en la expansión del acceso bajo estas características se ha dado a expensas de la calidad del servicio, como se advertirá en el siguiente capítulo de este informe.

La consideración antes señalada sobre las altas tasas de cobertura de la educación primaria, sin embargo, no debiera hacernos perder de vista el hecho de que aún existe un 4% de la población entre 6 y 11 años que no se encuentra asistiendo a ningún programa educativo y que necesita ser incorporado a este por medio de programas intersectoriales y focalizados, en tanto su situación social y familiar probablemente demanden una atención que va más allá de lo estrictamente educativo; sobre todo si consideramos que en el caso del área rural y en el grupo de los estudiantes que se encuentran en el grupo de pobres extremos este indicador se incrementa a 8%.⁴⁷

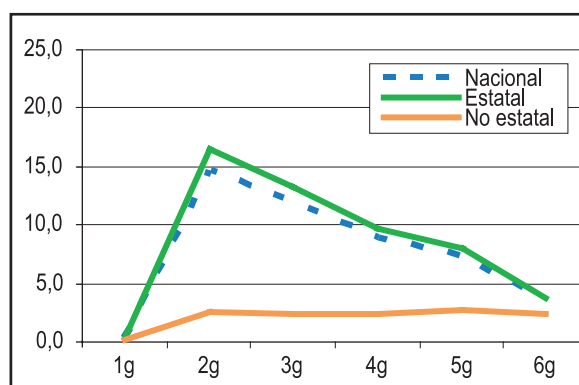
Ahora bien, se debe considerar que junto con la alta tasa de cobertura alcanzada para la educación primaria, aún se mantienen importantes tasas de desaprobación, las cuales, a su vez, producen elevadas tasas de atraso escolar que repercuten negativamente en la

46. Aproximadamente el 70% de las instituciones educativas primarias son multigrado o unidocentes y el 89% de las mismas se ubica en el área rural.

47. Estas cifras corresponden a la tasa de Cobertura Total para la población de 6 a 11 años y se pueden consultar en MED 2005b.

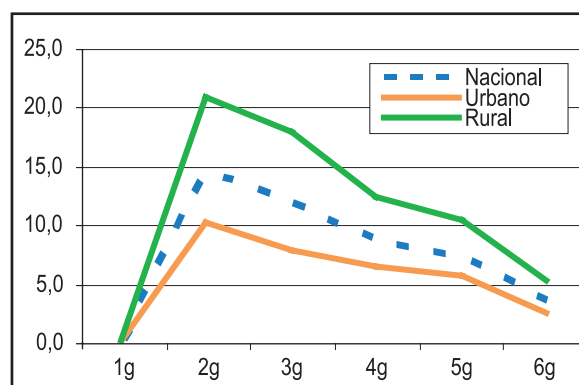
probabilidad de culminación de los estudios primarios.⁴⁸ Como se puede apreciar en los siguientes gráficos, las mayores tasas de desaprobación se presentan en los primeros grados, exceptuando al primer grado,⁴⁹ las cuales disminuyen, luego, en los dos últimos grados del nivel primario. Este hecho, si bien no es privativo de nuestro país, supone un problema serio en tanto los primeros grados de la educación primaria son el punto de entrada al sistema educativo, en que se construyen los fundamentos y los aprendizajes esenciales que condicionarán, positiva o negativamente, los futuros aprendizajes, la autoestima y la confianza de los niños (Torres 1995). De allí que sea imprescindible diseñar e implementar estrategias efectivas destinadas a reducir la repetición asociada al rendimiento inadecuado en esos grados fundamentales.

Gráfico 4.2: Tasa de desaprobación por grado según gestión de la IE



Fuente: UEE Censo Escolar 2003.
Elaboración: UMC.

Gráfico 4.3: Tasa de desaprobación por grado según área geográfica



Fuente: UEE Censo Escolar 2003.
Elaboración: UMC.

48. El atraso escolar se produce básicamente por ingreso tardío o por repetición. En nuestro país, al parecer el factor que incide más es el segundo. Según la información de las Encuestas Nacionales de Hogares (ENAH), la edad promedio de ingreso a la educación primaria ha mejorado significativamente en los últimos años. Así, en el año 2002, el 92% de los niños de seis años se encontraba cursando primero o segundo grado de primaria. Este porcentaje disminuía a 87,6 en el área rural. Junto con este aspecto, es preciso mencionar el hecho que cada vez se incrementa más la proporción de niños de cinco años que asiste al primer grado de educación primaria. Según la misma fuente de información, el 45,4% de los niños de cinco años está adelantado, es decir, a pesar de que por su edad no deberían estar asistiendo a primer grado, lo hacen.

49. En este grado existe promoción automática.

Si bien la edad promedio de ingreso a la educación primaria ha mejorado significativamente en los últimos años, todavía un 4,3% de 6 años lo hace con atraso, siendo este indicador mayor en las áreas rurales (6,2%) Este hecho probablemente se presenta de manera más acentuada en las zonas dispersas que presentan grandes distancias entre los hogares y las instituciones educativas.

Por otro lado, se debe considerar que tanto la entrada tardía a la educación primaria como la repetición tienen como resultado el atraso escolar o la extraedad. Este problema suscita que al interior de las aulas prevalezca la dispersión o heterogeneidad de edades de los alumnos, lo que dificulta el proceso de enseñanza/aprendizaje debido, en principio, a que diferentes niveles de madurez exigen diferentes estilos pedagógicos y a la necesidad de generar experiencias de aprendizaje que resulten interesantes para alumnos de toda la gama de edades. Ambas son tareas difíciles para las cuales pocos docentes están capacitados.

Otro de los factores que subvierte la probabilidad de la conclusión de la educación primaria está asociado con el retiro y la deserción del sistema educativo. Del mismo modo que la repetición, las tasas más altas de este indicador se encuentran en los primeros grados y presentan las mismas inequidades en razón a gestión y área de residencia.

Si bien diversos estudios han demostrado que la deserción está asociada en parte a la repetición, este problema es bastante más complejo debido a los múltiples factores sociales e individuales que influyen para que esta se produzca y frente a los cuales el sistema educativo no siempre tiene respuestas adecuadas y oportunas. No obstante, resulta primordial realizar investigaciones que permitan dar con los incentivos correctos que logren retener a los alumnos y disminuir la deserción al interior del sistema educativo. Muchos estudios señalan que esto es posible de lograr por medio de programas sensibles a la realidad local y flexibles para acoger a aquellos que tienen dificultades para permanecer en el sistema educativo.⁵⁰

Así, entonces, tanto la entrada tardía, la desaprobación, el retiro y la deserción ocasionan que la proporción de la población que concluye la educación primaria a la edad establecida normativamente sea aún bastante inferior al 100%. De esta manera, aproximadamente solo el 63% de la población de doce años —edad en la que se espera normativamente que el alumno ya se encuentre en el primer grado del nivel secundario— ha terminado el nivel primario.⁵¹ En otras palabras, solo este porcentaje ha concluido la educación primaria sin atrasarse.

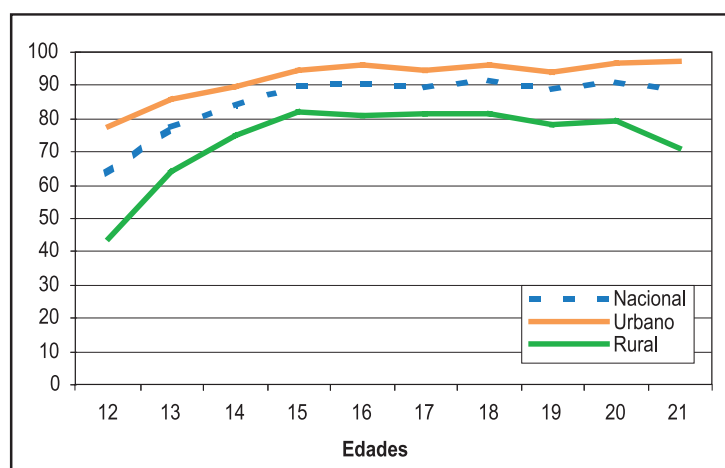
Sin embargo, es necesario señalar que esto no significa que solo este porcentaje completa este nivel, en tanto, como lo hemos mencionado anteriormente, la población escolar se atrasa en su proceso de adquisición de escolaridad y termina la primaria con una edad superior a la de la norma. Esto se puede observar en el gráfico 4.4 en el que se muestra la probabilidad de concluir el nivel primario a edades superiores a los doce años.

50. Uno de los factores que podría estar influyendo en la deserción de la población de seis a once años es la participación de dicha cohorte en la fuerza laboral. Así, la tasa de participación de la fuerza laboral de la población de seis a once años es de 19,3%, tasa que se desagrega en 5,4% y 37,4% para el área urbana y rural, respectivamente.

51. Este indicador se conoce como Índice Global de Escolarización (IGE) que es una suerte de medida resumen del éxito global del sistema ya que expresa la probabilidad real que tuvo una cohorte determinada de ingresar oportunamente al sistema educativo y mantenerse en él continuamente y sin atrasarse, es decir, da información sobre dos objetivos claves del sistema: el ingreso oportuno y el éxito escolar. Al respecto, véase: MED 2002 (p. 11-26).

Este representa la evolución de la probabilidad de concluir la primaria en edad por encima de lo esperado normativamente. Así, este índice va aumentando conforme se incrementan las edades hasta alcanzar su máximo valor a los 20 años con 91%, luego empieza a decrecer. Esto significaría que actualmente, a lo sumo este porcentaje de la población estaría culminando educación primaria en nuestro país. En otras palabras, todavía hay un 10% de la población que no ha logrado completar este nivel (MED, 2002).

Gráfico 4.4: Probabilidad de concluir la educación primaria según área geográfica



Fuente: MED 2002.
Elaboración: UMC.

Este indicador muestra brechas significativas por área de residencia. De este modo, en las áreas urbanas, el 77,5% de la población de doce años ha concluido la educación primaria frente al 43,9% en las áreas rurales. A pesar de que el valor del indicador mostrado en el anterior gráfico aumenta con la edad en ambos grupos de población, las brechas se mantienen a lo largo de las edades. Los datos sugieren que alrededor del 96% de la población urbana culmina la primaria mientras que solo un 81% lo hace en las áreas rurales.

Así, entonces, como se observa en el gráfico, el progreso hacia la conclusión de la educación primaria se hace cada vez más difícil a medida que se avanza hacia la meta, ya que ello supone atender a segmentos de la población que presentan mayores rasgos de marginación. Esto probablemente tenga que ver con su ubicación territorial (en zonas apartadas y de difícil acceso) o a la pertenencia de los niños y niñas a estratos sociales que enfrentan dificultades mucho mayores para progresar en este nivel, lo que se traduce en elevadas tasas de deserción y repetición. En consecuencia, las políticas tendientes a la universalización de la educación primaria no solo deben identificar a esos segmentos sino también adoptar estrategias especiales para garantizar su acceso al sistema educativo y retenerlos en este (Machinea, Bárcena y León 2005).

4.2. Educación secundaria: acceso, tránsito y conclusión

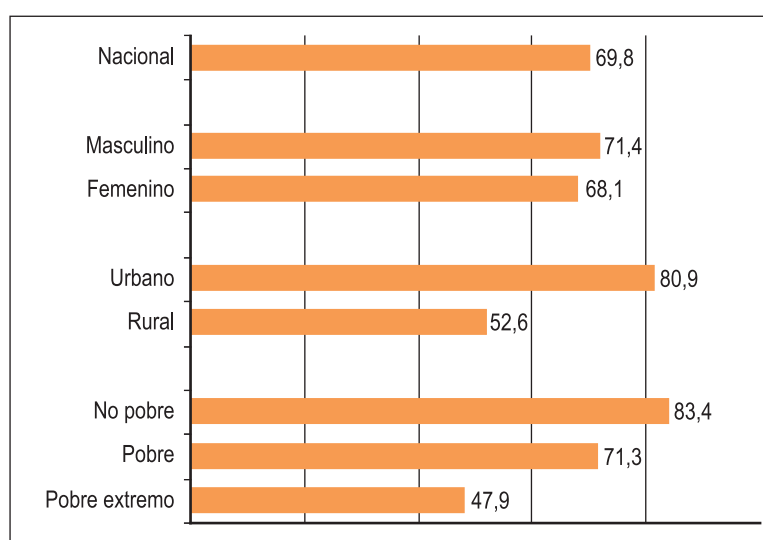
Si bien desde una perspectiva histórica se puede afirmar que la educación peruana ha alcanzado importantes progresos en los niveles de acceso al sistema por parte de las personas en edades de escolarización obligatoria, lo que se ha traducido en el incremento de los años de escolaridad, no deja de ser cierto, también, que actualmente los peruanos

en promedio no alcanzan a aprobar el tercer año de secundaria, en tanto los resultados de la última Encuesta Nacional de Hogares (2003) presenta un valor de 8,9 para este indicador.

Este indicador, nos permite constatar que nuestra población aún todavía no logra completar la educación básica obligatoria. En efecto, del grupo de edad de 12 a 16 años,⁵² que aproximadamente constituye el 10,6% de la población nacional, tan solo el 85,8% asiste a algún centro del sistema educativo. Si dividimos a esta población según el área geográfica tenemos que 86,5% en el área urbana asiste a alguna educación educativa mientras que el 77% lo hace en el área rural; en esta última todavía se observan diferencias por sexo (83,2 hombres vs. 72,6% mujeres)

Este déficit en el acceso de dicha cohorte de edad tiene que ver en parte con la limitada oferta de educación secundaria, actualmente a cargo de 9 532⁵³ instituciones educativas concentradas básicamente en las zonas urbanas y centros poblados mayores (62,9%)⁵⁴, así como con los costos de oportunidad que tiene este grupo de edad debido a su participación en la fuerza laboral.⁵⁵

Grafico 4.5: Tasa de cobertura en nivel de la población de 12 a 16 años, según sexo, área y nivel de pobreza



Fuente: MED (2005b).
Elaboración: UMC.

52. Edades normativas correspondientes al nivel secundario.

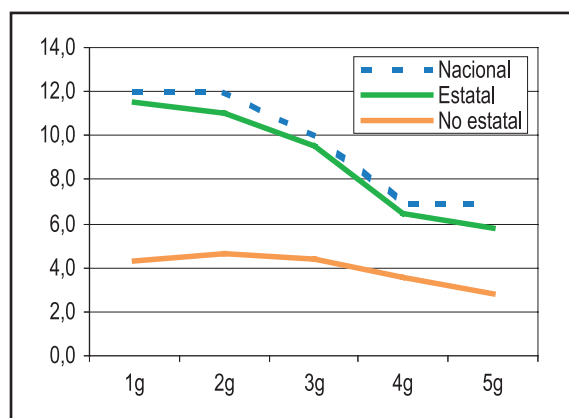
53. Frente a las 33 462 instituciones educativas primarias.

54. Si bien en los años noventa el crecimiento de la matrícula secundaria se ha dado con mayor acento en el área rural, su incremento en los próximos años está sujeto al desarrollo de una política de expansión de la oferta educativa de secundaria en dicho ámbito, lo que dista de ser un proceso sencillo en el marco de los limitados recursos presupuestales con los que cuenta y que, además, exige pensar en una forma de administración de los servicios educativos diferente a la vigente, la que es difícil de reproducir en contextos de escasa población. En efecto, si la demanda de las zonas rurales más alejadas para el caso del nivel primario se pudo cubrir a través de los CE unidocentes y multigrado, el carácter especializado de las diferentes materias del nivel secundario torna inviable esta alternativa. En este sentido, asegurar la atención a una población asentada sobre un patrón muy disperso exige estrategias creativas, como la educación a distancia y la modalidad de alternancia, experiencias que el Ministerio y la sociedad civil vienen implementando.

55. A nivel nacional, el 33,1% de la población de esa edad trabaja además de estudiar. En la zona urbana este porcentaje corresponde al 16,5% mientras que en el área rural es de 59,3%.

Si bien, lo señalado líneas arriba sobre la población de 12 a 16 años ya muestra cierto déficit en la asistencia al sistema educativo, la situación en que se encuentra dentro de este exhibe aún mayores problemas. Así, solo el 69,8% de dicho grupo de edad asiste al nivel que normativamente le corresponde (secundaria) y aproximadamente el 46% en promedio lo hace al grado que le corresponde según su edad.

Gráfico 4.6: Tasa de desaprobación por grado según gestión de la IE

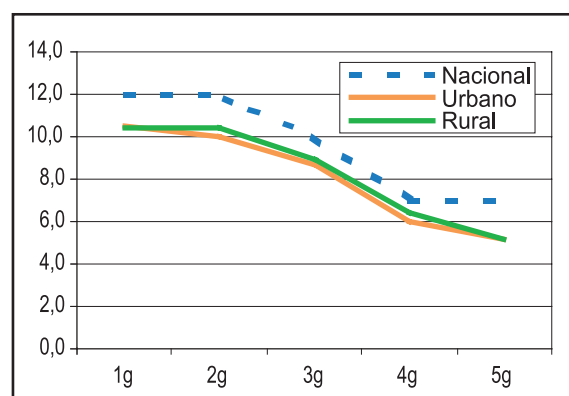


Fuente: UEE Censo Escolar 2003.
Elaboración: UMC.

De otro lado, la información muestra que las tasas de asistencia de la población de 12 a 16 años que asiste a una escuela secundaria de acuerdo al sexo siguen una clara tendencia a equipararse, aunque todavía mantienen una ligera ventaja a favor de los hombres.

Uno de los factores que explica el hecho de que un gran porcentaje de adolescentes y jóvenes de 12 a 16 años no se encuentre cursando los estudios en el nivel ni en el grado que les corresponde está asociado en gran parte con la experiencia de la desaprobación de algún grado y/o el abandono temporal o permanente del sistema educativo. Evidentemente, en el caso de la educación secundaria dicho atraso se acarrea desde el nivel primario, en el cual, como se ha podido observar en la sección anterior, los niveles de desaprobación y retiro son bastante significativos.⁵⁵

Gráfico 4.7: Tasa de desaprobación por grado según área geográfica



Fuente: UEE Censo Escolar 2003.
Elaboración: UMC.

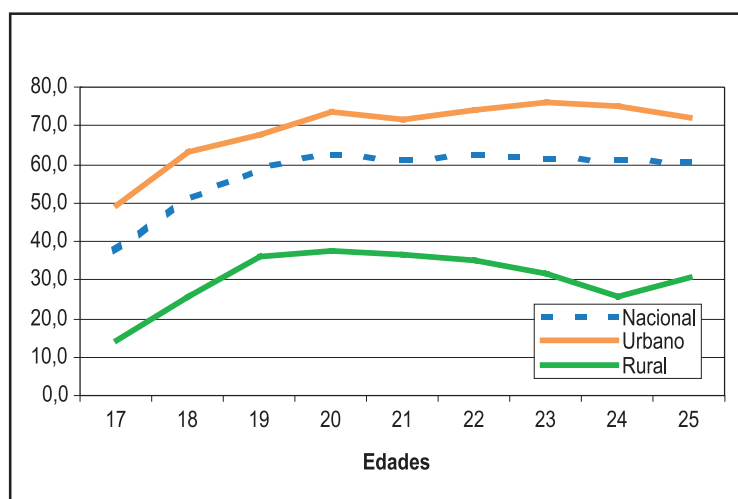
55. En efecto, la extraedad o el atraso que tiene un estudiante a lo largo del ciclo escolar solo puede mantenerse o aumentar, pero de ninguna manera disminuir. Al respecto véase «Cobertura y Escolarización» en Guadalupe 2002.

Como se puede observar en los gráficos anteriores, las mayores tasas de desaprobación se presentan en los primeros grados y van disminuyendo en forma constante mientras se va avanzando en los grados de este nivel. Además, se puede observar que hay una disparidad respecto a este indicador entre las escuelas gestionados por el Estado de aquellas que no lo son. Mientras que las tasas de desaprobación para las escuelas rurales y urbanas son bastante similares, situación bastante diferente de lo observado en el caso del nivel primario. En las tasas de retiro sí se observa una brecha importante entre las escuelas urbanas y rurales (5% y 11% respectivamente) a favor de las primeras.

Los indicadores mostrados revelan el atraso escolar en el grupo de edad de 12 a 16 años, lo que se traduciría en que aproximadamente solo el 37% de la población de 17 años concluya la secundaria. Es decir únicamente este porcentaje logra concluir la educación básica en los 11 años, tal como el sistema espera que lo haga la totalidad de los estudiantes.

Sin embargo, al igual que en el nivel primario, es necesario precisar que esto no significa que solo este porcentaje completa este nivel, en tanto, como lo hemos mencionado anteriormente, la población escolar se atrasa en su proceso de adquisición de escolaridad y termina la secundaria con una edad superior a la de la norma. Esto se puede observar en el siguiente gráfico, en el que se muestra el indicador de probabilidad de conclusión de este nivel para edades superiores a los 17 años. Este representa la evolución de la probabilidad de concluir la secundaria en edad por encima de lo esperado normativamente. Así, este índice va aumentando conforme se incrementan las edades, se eleva hasta valores que oscilan entre 61 y 62% entre los 20 y 24 años. Esto estaría significando que a lo sumo el 62% de la población está culminando la secundaria y, en consecuencia cerca del 38% de la población no concluye la educación básica (MED, 2002).

Gráfico 4.8: Probabilidad de concluir la educación secundaria según área geográfica



Fuente: MED 2002.

Asimismo, este indicador revela una fuerte brecha por área geográfica. Así, mientras el 49,3 % de la población de 17 años en el área urbana ha concluido la educación secundaria, en el área rural solo lo ha hecho el 14,1%.

Diversos estudios han señalado la necesidad de establecer una meta con respecto al incremento del término de la educación secundaria. Esta meta se sustenta en la evidencia de que la conclusión de este nivel constituye actualmente un umbral educativo que eleva significativamente la probabilidad de mantenerse fuera de la pobreza absoluta en la vida activa, toda vez que se traduce en un incremento importante de los ingresos laborales. Hace ya una década, la CEPAL estimó que extender la escolaridad secundaria es socialmente más rentable que compensar posteriormente la falta de educación con programas de capacitación de adultos. Los programas compensatorios de educación de adultos cuestan de 1,5 a 5 veces el valor de cuatro años de enseñanza secundaria, lo que muestra las ventajas de invertir oportunamente en este nivel (Machinea, Bárcena y León 2005).



5.1. Resultados nacionales según niveles de desempeño

En el presente capítulo se presentan los resultados nacionales que obtuvieron los estudiantes en la EN 2004. Dichos resultados, tal como fuera señalado en el capítulo 2, se expresan en términos de niveles de desempeño que dan cuenta del grado de desarrollo de las capacidades establecidas en el currículo.⁵⁶

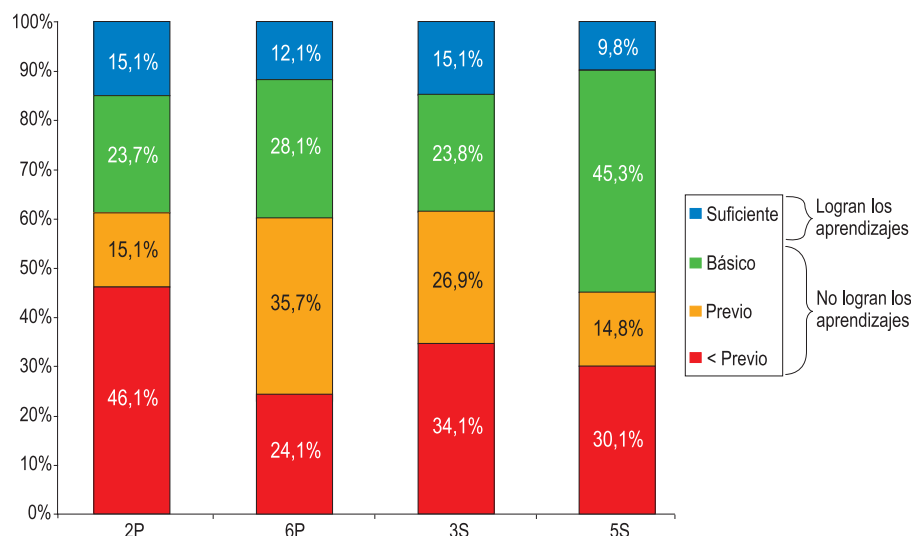
En el capítulo señalado se mencionó que son los estudiantes ubicados en el nivel suficiente los que demuestran haber alcanzado los aprendizajes esperados para el grado que cursan al menos en las capacidades evaluadas en las pruebas.

Asimismo, se señaló que el diseño de la evaluación permite establecer no solo el nivel de desempeño esperado de las capacidades para el grado evaluado (nivel suficiente), sino también permite definir niveles de desempeño que denotan un desarrollo parcial de dichas capacidades. Así, el estudiante ubicado en el nivel básico demuestra un desarrollo incipiente o inicial de las capacidades del grado, mientras que el estudiante ubicado en el nivel previo demuestra un desarrollo de capacidades propio de grados anteriores al que cursa. Finalmente, se ha establecido, por defecto, un grupo denominado *debajo del previo* en el que se ubican los estudiantes que no logran siquiera realizar todas las tareas correspondientes al nivel previo.

En el gráfico 5.1, que se presenta a continuación, se muestran los resultados nacionales en Comprensión de textos escritos en los cuatro grados evaluados. Dichos resultados están expresados en términos de porcentajes de estudiantes ubicados en cada uno de los niveles de desempeño mencionados.

56. Se recomienda que al leer esta sección del reporte se revise la descripción de los niveles de desempeño presentada en la sección final del capítulo 2 con el fin de relacionar los porcentajes de los estudiantes en cada uno de los niveles de desempeño con las tareas que estos son capaces de resolver en cada uno de los mismos.

Gráfico 5.1: Nacional: Porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño en Comprensión de textos escritos



El gráfico muestra que los porcentajes de los estudiantes ubicados en el nivel suficiente para los cuatro grados evaluados son sumamente bajos considerando que, pese a que este nivel es el esperado para todos los estudiantes, en ningún caso este porcentaje excede al 16%. Es el quinto grado de secundaria el que presenta el menor porcentaje de estudiantes en este nivel (9,8%).

Si bien en términos generales resultan preocupantes estos resultados, es importante llamar la atención sobre el bajo nivel de logro alcanzado por los estudiantes de segundo grado. De hecho, es la primera vez que se cuenta con evidencia empírica a escala nacional sobre el nivel de logro de los estudiantes respecto a la comprensión de textos escritos al inicio de la educación primaria.⁵⁷

Como se puede observar en el gráfico 5.1, solo el 15% de los estudiantes de segundo grado logra los aprendizajes esperados en comprensión de textos. Esto significa que el 85% de los estudiantes de segundo de primaria no alcanza las metas propuestas para el grado de acuerdo con el currículo vigente. Asimismo, de este gráfico se desprende el hecho de que más de la mitad de estudiantes de este grupo no llega siquiera a mostrar el desarrollo de las capacidades establecidas para el primer grado lo que evidencia una situación en extremo preocupante dado que los primeros años de escolaridad son los que sientan las bases para los aprendizajes posteriores.⁵⁸ Esta evidencia abona a la creciente demanda que se está haciendo desde diversos sectores para dar prioridad a la educación inicial.⁵⁹

57. Uno de los grandes retos que asumió la EN 2004 fue la evaluación del segundo grado de primaria con las características que comporta una evaluación estandarizada con instrumentos de lápiz y papel. Evaluar a estudiantes de este grado supuso un diseño particular de las pruebas y un tratamiento especial de la aplicación de instrumentos, sobre todo porque es en los primeros grados en los que se afianzan las habilidades básicas de lectura y escritura.

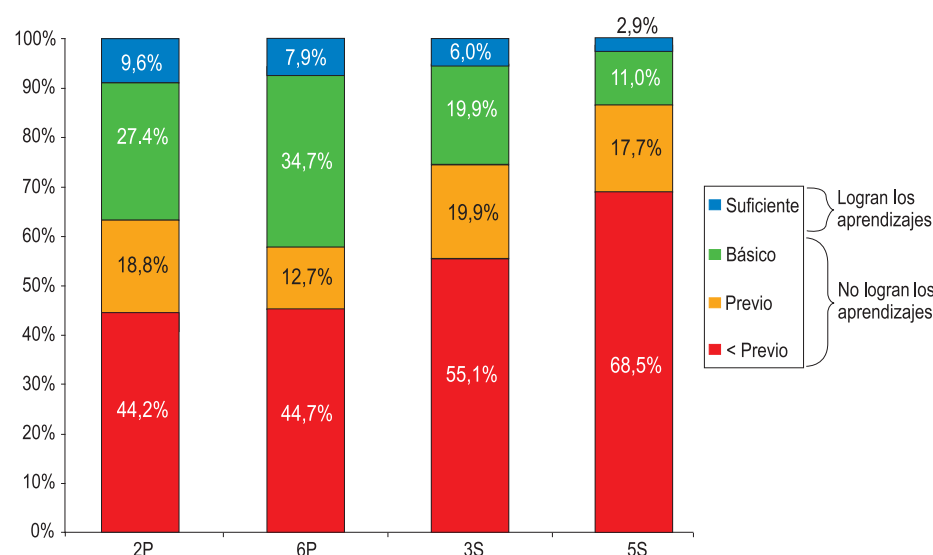
58. La investigación realizada por Cueto (2004) desde un enfoque longitudinal evidencia que el rendimiento previo es la variable que tiene más probabilidad de asociarse con los aprendizajes posteriores.

59. El Proyecto Educativo Nacional presentado por el Consejo Nacional de Educación propone como una política específica del primer objetivo estratégico «Oportunidades y resultados educativos de igual cali-

En efecto, la mayoría de los estudiantes de segundo grado solo muestra evidencias de poder leer palabras aisladas o leer oraciones sueltas fuera de contexto (UMC 2005a). Actualmente, sabemos que la alfabetización no termina en los dos primeros años de la primaria, cuando los estudiantes tienen el conocimiento básico de las herramientas de lectura y escritura, en tanto no se trata de una capacidad que los estudiantes posean o no posean, sino que es un proceso que se desarrolla a lo largo de nuestras vidas a partir del contacto con diversa información escrita. De allí que resulta fundamental que los estudiantes de este grado logren consolidar el aprendizaje básico de la lectura y la escritura de manera que les permita continuar con su proceso de alfabetización y, con ello, seguir aprendiendo y desarrollando habilidades en otras áreas y grados durante su vida escolar y en otros ámbitos de la vida en sociedad.

En el gráfico 5.2 que se presenta a continuación se muestran los resultados en el área de Matemática en los cuatro grados evaluados.

Gráfico 5.2: Nacional: Porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño en Matemática



Al igual que en el caso de Comprensión de textos escritos los porcentajes de estudiantes ubicados en el nivel suficiente en los cuatro grados evaluados son bastante bajos e incluso se observa una clara tendencia al descenso conforme se avanza en la educación básica. De hecho, a diferencia de lo que ocurre en Comprensión de textos escritos, en el caso del área de Matemática se presenta una tendencia más clara al incremento del porcentaje de estudiantes ubicados en el grupo por debajo del nivel previo. En el caso de quinto de secundaria este porcentaje llega a 68,5%. Estos resultados parecen indicar, aún cuando se trata solo de una hipótesis,⁶⁰ que en esta área los años adicionales de escolaridad no están contribuyendo a ampliar y desarrollar las habilidades matemáticas de los estudiantes.

dad para todos», la universalización de la educación inicial de niños de 4 y 5 años. Esta política propone garantizar una educación inicial de calidad y el logro efectivo de aprendizajes fundamentales para todos los niños y niñas de 4 y 5 años de edad (CNE 2005).

60. En la EN 2004 no se ha aplicado un diseño de equiparación vertical. Es decir, los resultados de los cuatro grados evaluados no se encuentran en la misma escala. Sin embargo, considerando la información de los puntos de corte se evidencia que cada vez hay un mayor número de estudiantes debajo del nivel previo.

Si bien el porcentaje de estudiantes que se ubica en el nivel esperado (nivel suficiente) en Matemática al inicio de la primaria es tan bajo como en Comprensión de textos escritos, los resultados que llaman más la atención en esta área son los correspondientes a los estudiantes de quinto grado de secundaria. Es decir, de los jóvenes que están próximos a concluir su educación básica, solo el 2,9% demuestra haber desarrollado las habilidades y destrezas matemáticas y haber consolidado las nociones y los contenidos que se esperan de ellos, considerando los objetivos propuestos por la estructura curricular en dicha área.⁶¹

Como se puede observar en el gráfico 5.2, de la proporción de estudiantes que no logra alcanzar los aprendizajes esperados, cerca del 70% de los estudiantes de quinto de secundaria se ubica en el grupo por debajo del previo. Esto significa que aproximadamente siete de cada diez estudiantes que están terminando su educación secundaria solo son capaces de resolver tareas como situaciones problemáticas sencillas y rutinarias de tipo comercial que demandan aplicar una secuencia de operaciones aritméticas básicas (adición, sustracción, multiplicación o división); identificar formas de figuras geométricas elementales y sus principales elementos, como ángulos y lados; y calcular el promedio de un conjunto pequeño de números naturales.⁶² Como se puede apreciar, estas tareas son bastante elementales y se encuentran lejos de lo que debería ser capaz de desarrollar un estudiante en el área de Matemática luego de, al menos, once años de escolaridad. En este sentido, es bastante probable que estos estudiantes tengan serias dificultades para emplear la matemática como una herramienta significativa y eficiente en su vida; posiblemente sus dificultades serán aún mayores para enfrentar situaciones problemáticas novedosas propias de la vida cotidiana y para ampliar sus conocimientos y seguir desarrollando sus capacidades en esta y en otras áreas, en caso de que decidan continuar su educación formal.

Adicionalmente a lo señalado, un aspecto sobre el cual consideramos importante llamar la atención es el grado de variabilidad que presenta el rendimiento de los estudiantes en ambas áreas,⁶³ dada las implicancias pedagógicas que conlleva esta constatación para el trabajo en aula. Estas implicancias no siempre son lo suficientemente subrayadas a pesar de la importancia que conllevan. En efecto, si bien es cierto que un docente debe estar en capacidad de generar estrategias de enseñanza adecuadas a los niveles de desarrollo de sus estudiantes, el hecho de que un docente tenga a su cargo un aula en la que prevalezca una alta heterogeneidad en dichos niveles podría dificultar seriamente su labor si este no cuenta con las herramientas necesarias que le permitan, considerando esa diversidad, aprovechar las diferencias para generar aprendizajes significativos en cada uno de sus estudiantes. En tal sentido, es necesario que la política educativa —en particular la concerniente a la formación y capacitación docente— atienda de manera pertinente a este problema.

61. Esta información resulta más preocupante aún si se considera el hecho de que los estudiantes de este grado constituyen una población seleccionada de aquellos que ingresan al primer grado de primaria. En un estudio realizado se estimó que de cada 1 000 niños que inician sus estudios primarios sólo cerca de la mitad (520) logran culminar la secundaria (Guadalupe, César y Liliana Miranda «Eficiencia interna» en: Guadalupe 2002).

62. Al respecto consultar: Informe Pedagógico de Resultados. Formación Matemática. Tercer Grado de Secundaria. Quinto grado de Secundaria (UMC 2005d).

63. La elaboración de los informes de resultados para cada escuela evaluada, que ha realizado por primera vez la UMC, nos ha permitido observar que la mayor parte de las aulas evaluadas tienden a presentar una alta heterogeneidad en el rendimiento de los estudiantes.

De otro lado, resulta relevante observar si existe relación entre el rendimiento de los estudiantes en Matemática y en Comprensión de textos escritos. Los hallazgos de la EN 2004 corroboran lo encontrado en evaluaciones anteriores (UMC 2004:95)⁶⁴ sobre la existencia de una relación positiva en el rendimiento en ambas pruebas, es decir que los estudiantes que mejor rinden en Matemática tienden a hacerlo de igual manera en Comprensión de textos escritos y viceversa.

Cuadro 5.1: Relación entre las medidas del rendimiento en Matemática y Comprensión de textos escritos

Grado	Correlación
Segundo de primaria	0,78
Sexto de primaria	0,68
Tercero de secundaria	0,61
Quinto de secundaria	0,54

Como se puede advertir en el cuadro 5.1, en todos los casos los coeficientes de correlación calculados han resultado significativos y altos. Se observa que en general las relaciones entre el rendimiento en Matemática y Comprensión de textos son más fuertes en los grados de primaria, frente a lo que sucede en secundaria. Es decir, conforme se avanza en el ciclo educativo, la correlación es menos fuerte. Esto estaría indicándonos que los rendimientos en ambas áreas van diferenciándose cada vez más, a pesar de mantener un nivel moderado de asociación.

¿Cómo explicar la disminución de la relación directa del rendimiento en ambas áreas? Este tema resulta bastante complejo de ser abordado y ameritaría una investigación específica sobre el mismo; sin embargo, a modo de hipótesis general podríamos señalar que, en el nivel secundario, el área de Matemática se hace una disciplina más especializada, que demanda de parte del estudiante tanto el manejo de una mayor diversidad de estrategias como de un mayor desarrollo de nociones y contenidos, razón por la cual la comprensión de la situación problemática por medio de la adecuada lectura es una condición necesaria, pero no suficiente por sí misma para una adecuada elaboración de la estrategia de resolución. En los grados superiores los problemas se van complejizando cada vez más, lo cual constituye un factor de dificultad que opera en mayor grado que la comprensión lectora de los estudiantes.

5.2. Resultados de los estudiantes según diversos ejes de equidad

Si bien, como se ha visto en el acápite anterior, los resultados de la EN 2004 muestran un problema general en la calidad de los aprendizajes en todos los estratos estudiados, también se evidencia la existencia de brechas de equidad en el desempeño de los estudiantes al interior del sistema educativo, tanto en el área de Matemática como en Comprensión de textos escritos.

64. En otros países también se ha encontrado este tipo de relaciones. Por ejemplo, Carvajal, Trejos y Soto (2004) han encontrado relaciones significativas entre estas áreas para el caso de las pruebas ICFES aplicadas en Colombia.

5.2.1. RESULTADOS DE LOS ESTUDIANTES SEGÚN LOS ESTRATOS REPRESENTATIVOS DE LA MUESTRA

Como se puede apreciar en los siguientes cuadros, los resultados de la EN 2004 muestran la existencia de brechas entre diferentes grupos evaluados de estudiantes. Como se ha señalado en el capítulo 2, los estratos representativos en la EN 2004 fueron, además del nacional, el tipo de gestión de la IE evaluada (estatal / no estatal), el área geográfica donde se ubica la IE estatal (urbana / rural) y la característica de la IE estatal (polidocente/multigrado).

Los estudiantes de las IE estatales muestran un desempeño menor en comparación con los estudiantes de las IE no estatales en todos los grados evaluados. Entre los estudiantes de las IE estatales, se observa un mejor desempeño de los estudiantes del área urbana en comparación a los estudiantes de zona rural.

En primaria, los estudiantes de las IE estatales polidocentes obtienen mejores resultados que sus pares en las IE estatales multigrado. Según sexo, los hombres muestran mejor desempeño que las mujeres en el área de Matemática en todos los grados evaluados. En la prueba de Comprensión de textos escritos del área de Comunicación, a pesar de que no se observan diferencias estadísticamente significativas en segundo grado, en el resto de grados evaluados las mujeres muestran un mejor desempeño que los hombres.

Cuadro 5.2: Nacional: Porcentaje de estudiantes de los grados evaluados de educación primaria que alcanza el nivel suficiente

Grado y Área	Nacional	Nacional		Tipo de gestión		IE Estatales		IE Estatales	
		Hombre	Mujer	Estatal	No estatal	Urbano	Rural	Polidocente	Multigrado
Segundo									
Comprensión de textos	15,1%	15,1%	15,2%	9,5%	48,5%	14,5%	2,5%	13,9%	1,7%
Matemática	9,6%	10,9%	8,3%	6,1%	30,4%	8,8%	2,4%	8,4%	2,2%
Sexto									
Comprensión de textos	12,1%	11,0%	13,3%	8,2%	36,1%	11,0%	2,3%	10,5%	1,7%
Matemática	7,9%	9,0%	6,8%	4,4%	29,7%	5,8%	1,3%	5,6%	0,9%

Cuadro 5.3: Nacional: Porcentaje de estudiantes de los grados evaluados de educación secundaria que alcanza el nivel suficiente

Grado y Área	Nacional	Nacional		Tipo de gestión		IE Estatales	
		Hombre	Mujer	Estatal	No estatal	Urbano	Rural
Tercero							
Comprensión de textos	15,1%	13,4%	16,7%	10,1%	34,1%	11,1%	3,3%
Matemática	6,0%	6,7%	5,2%	2,2%	20,7%	2,4%	0,2%
Quinto							
Comprensión de textos	9,8%	8,5%	11,0%	7,0%	20,0%	7,6%	2,1%
Matemática	2,9%	3,5%	2,4%	0,8%	10,5%	0,9%	0,0%

Si bien la muestra de la EN 2004 no fue representativa a escala departamental, con el fin de tener información referencial sobre el desempeño de los estudiantes en las distintas regiones del país, se han elaborado los siguientes gráficos que muestran la comparación del rendimiento según departamento, tomando la información correspondiente a sexto grado de primaria.

Gráfico 5.3: Sexto grado de primaria: comparación del rendimiento promedio en Comprensión de textos escritos según departamento

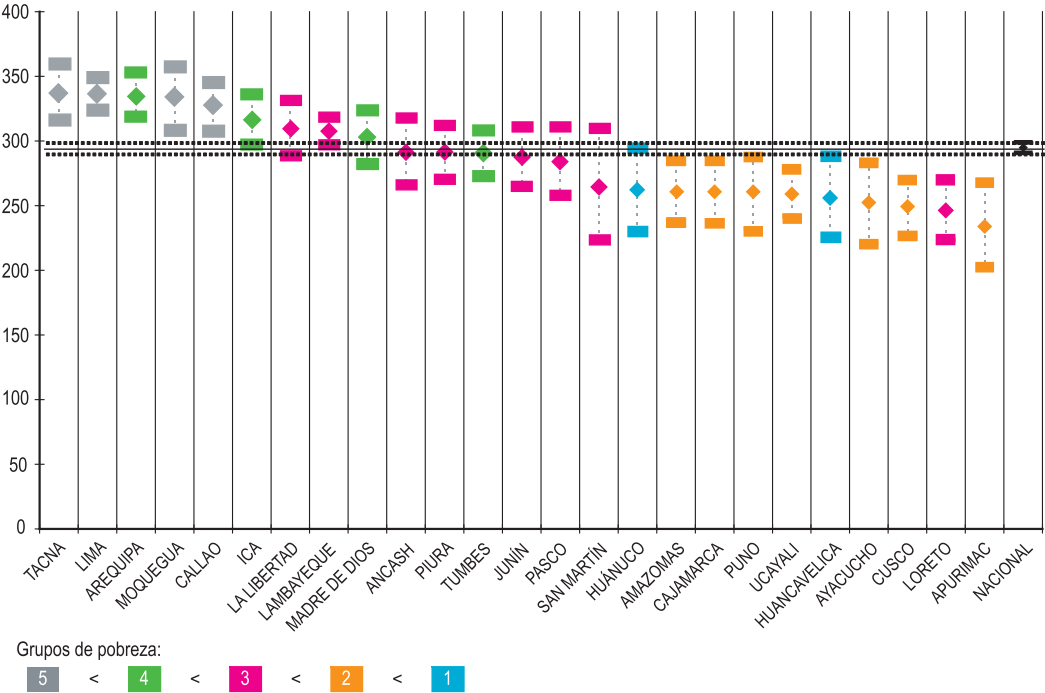
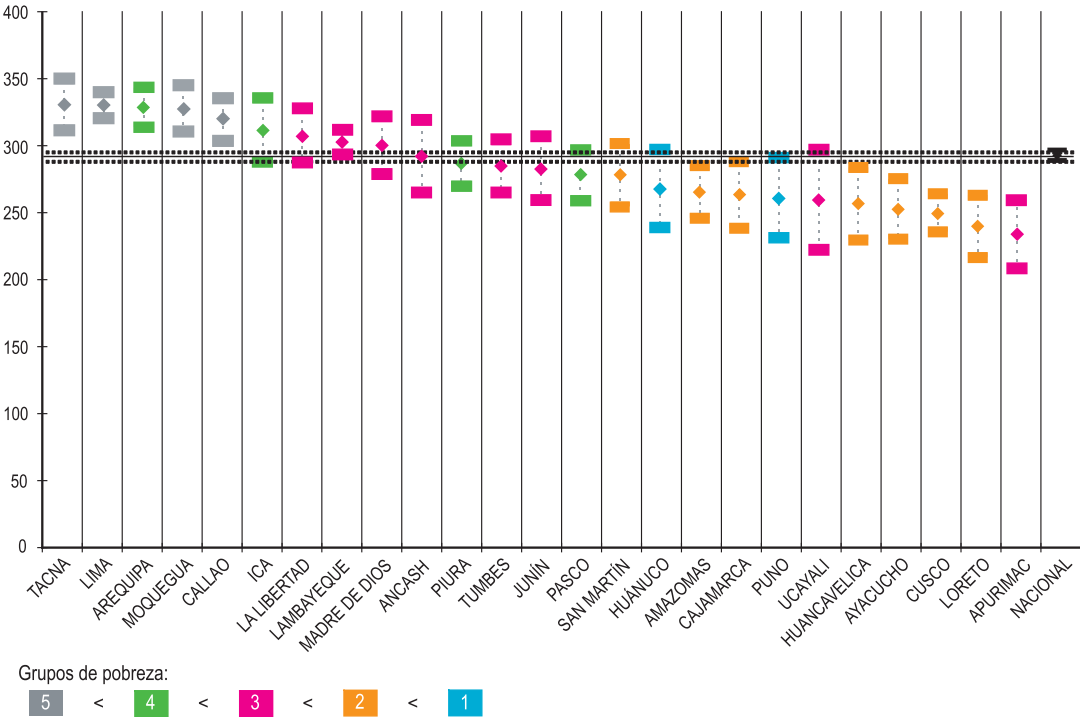


Gráfico 5.4: Sexto grado de primaria: comparación del rendimiento promedio en Matemática según departamento



Como se puede observar en ambos gráficos, los estudiantes que viven en los departamentos con menores niveles de pobreza⁶⁵ alcanzan mejores rendimientos en ambas áreas. Es el caso de los departamentos ubicados en la costa sur: Tacna, Lima, Arequipa, Moquegua, Ica, etc. En el otro extremo, con menores rendimientos, se encuentran los departamentos con mayores niveles de pobreza que son más rurales y con presencia de poblaciones cuya lengua materna es vernácula como, por ejemplo, Apurímac, Loreto, Cusco, Ayacucho, etc. Este ordenamiento supone que los resultados escolares acompañan también la jerarquía en la desigualdad socioeconómica de las regiones del país, lo que sugiere, como lo han señalado varios estudios,⁶⁶ que el gasto público no se asigna y distribuye adecuadamente de tal manera que las poblaciones más necesitadas reciban mayores recursos a fin de que la educación cumpla su objetivo de constituirse en un sistema de igualdad de oportunidades.

5.2.2. RESULTADOS DE LOS ESTUDIANTES DE SEXTO GRADO DE PRIMARIA POR NIVELES DE DESEMPEÑO SEGÚN DIVERSOS EJES DE EQUIDAD

Con el fin de que las diferencias en el rendimiento de los distintos grupos de estudiantes evaluados puedan observarse con más detalle, en esta sección se presentan los resultados de los estudiantes de sexto grado por nivel de desempeño.⁶⁷ Dichos resultados son presentados conjuntamente con información sobre algunas características socioculturales de los estudiantes pertenecientes a dichos estratos con el fin de que, al momento de leer los resultados, se tenga en consideración el contexto en que se producen los mismos. En efecto, si bien el análisis de factores asociados a rendimiento de los estudiantes se presentará en un próximo informe, se ha considerado importante incluir esta información en tanto, como señalan numerosos estudios (Benavides 2002; OECD 2003; UMC 2004), las características socioeconómicas y culturales de las familias de las que provienen los estudiantes tienden a estar fuertemente asociadas a los resultados de aprendizaje que estos consiguen en la escuela.⁶⁸

Al dividir la muestra nacional según el tipo de gestión de la institución educativa, se puede apreciar, tanto para el caso de Comprensión de textos como de Matemática, que el porcentaje de estudiantes que alcanza el nivel suficiente es mayor entre las IE no estatales. Es decir, los resultados muestran una diferencia significativa entre los estudiantes que estudian en escuelas estatales y aquellos que lo hacen en las no estatales. A pesar de ello, también es importante advertir que el porcentaje de estudiantes de las IE no estatales que logra un nivel de dominio adecuado no es muy elevado en tanto, en el mejor de los casos, no supera el 36%.

Otra diferencia importante entre los estudiantes pertenecientes a ambos estratos, es que en el caso de las instituciones de gestión estatal, la mayoría de alumnos se ubica en el nivel previo y por debajo de este, mientras que en las de gestión estatal, la mayoría

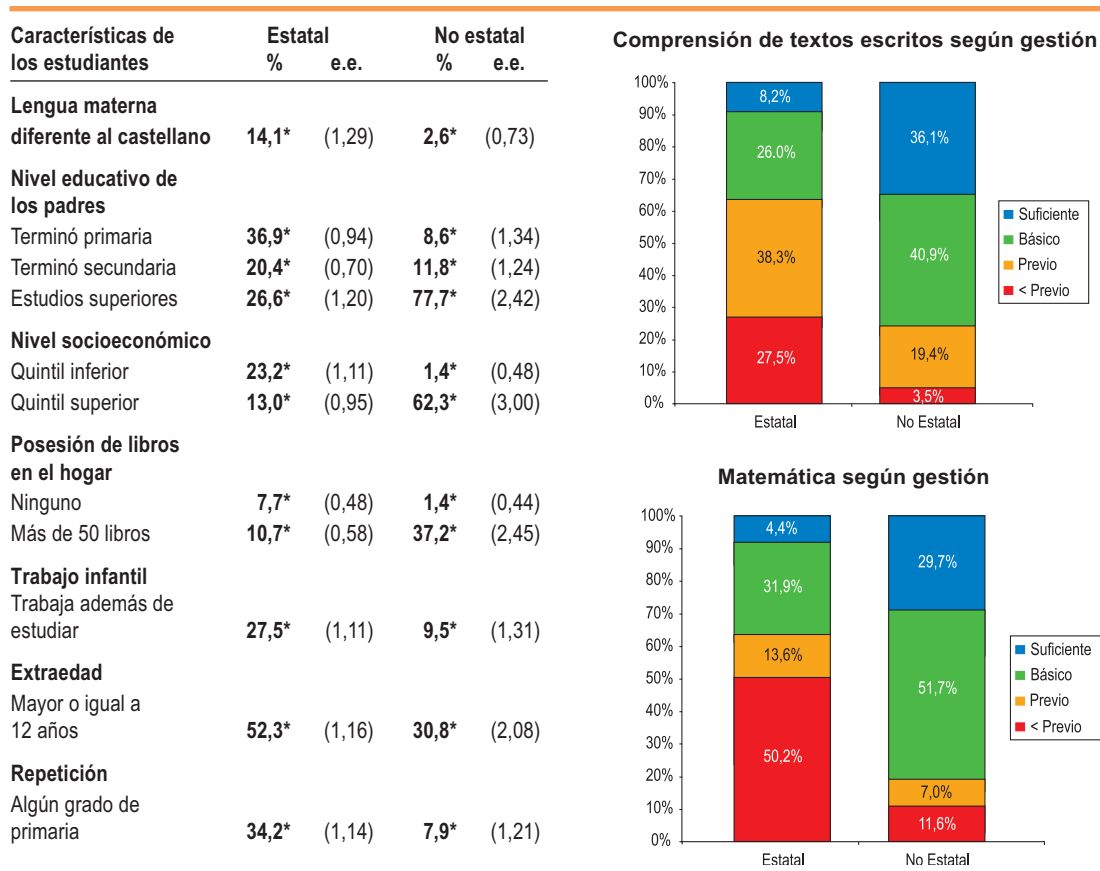
65. El ranking de pobreza por departamento se ha tomado de Herrera (2002).

66. Al respecto ver: Jaramillo y Arteaga (2003) y Saavedra y Suárez (2002).

67. Los resultados para los otros grados evaluados, pueden ser consultados en el anexo 5.

68. Sin embargo, ello no significa que las diferencias en el rendimiento se expliquen únicamente por variables familiares o individuales de los alumnos. De hecho la literatura sobre escuelas eficaces presenta evidencia sobre escuelas que en contextos socioeconómicos similares obtienen resultados diferentes lo que implica que los procesos pedagógicos e institucionales que se dan al interior de ellas inciden sobre los resultados de las mismas.

Cuadro 5.4: Sexto grado de primaria: porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño según la gestión de la IE



e.e.: error estándar.

* Diferencia estadísticamente significativa al 5%.

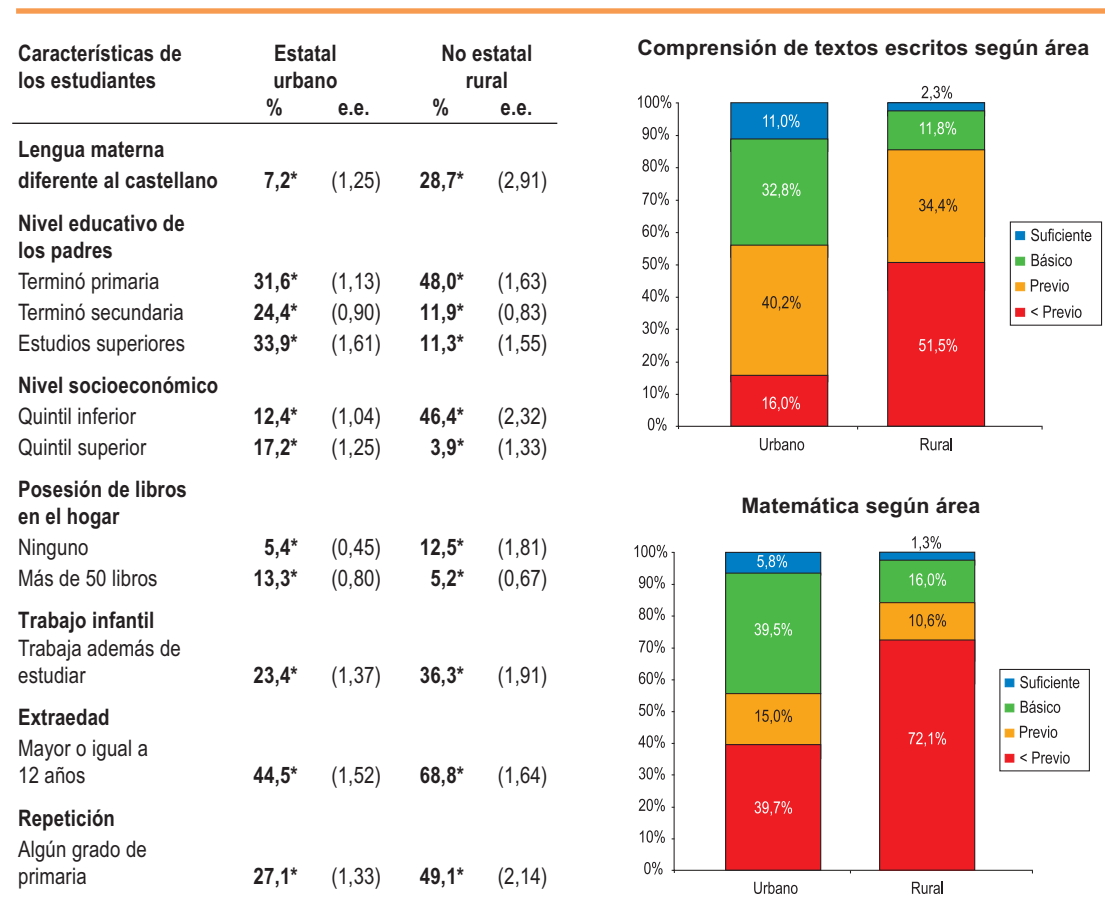
tiende a ubicarse en los niveles suficiente y básico. Ello implica que el desafío para remontar estos resultados es mayor en el caso de las IE de gestión estatal en tanto los estudiantes que asisten a este tipo de escuelas solo han desarrollado, en el mejor de los casos, un nivel de dominio relativo a grados anteriores al que actualmente cursan.

Sin embargo, estas diferencias en los rendimientos deben ser entendidas teniendo en cuenta algunas características familiares y escolares que las contextualizan. Como es ampliamente conocido, en nuestro país, los estudiantes provenientes de familias económicamente favorecidas son principalmente atendidos en las escuelas no estatales, mientras que los estudiantes provenientes de los estratos socioeconómicos más desfavorecidos asisten, en su mayoría a escuelas estatales. Esta distribución de la población estudiantil, entonces, debe ser considerada en la interpretación de las diferencias entre ambos estratos. Así, de acuerdo con lo mostrado en el cuadro 5.4, se puede observar que los alumnos pertenecientes a los colegios de gestión no estatal se caracterizan por contar con un conjunto de condiciones que —según diversas investigaciones— se encuentran asociadas positivamente con el rendimiento académico (OECD 2003; UMC 2004). El nivel educativo de los padres suele ser mayor y pertenecen a estratos socioeconómicos más altos. Estas características, entre otras, les permiten a estos estudiantes contar, por ejemplo, con un mayor número de libros en casa y con un soporte familiar, que constituye un capital cultural importante y que influye positivamente en sus rendimientos. En general, se aprecian también menores índices de repetición, extraedad y un menor porcentaje de estudiantes que trabajan.

Al interior de las escuelas estatales, el diseño muestral de la EN 2004 también permite subdividir este grupo en algunos estratos. Una de estas subdivisiones responde a la ubicación geográfica de la IE a la que asisten los estudiantes. Al analizar las diferencias de rendimiento en Comprensión de textos escritos y Matemática, se observa que existen diferencias significativas entre los ámbitos rural y urbano a favor de este último. Como se puede observar en el cuadro 5.5, el mayor porcentaje de estudiantes que logra alcanzar un nivel adecuado de dominio en las áreas evaluadas proviene de escuelas ubicadas en áreas urbanas.

Estas diferencias —que señalan el grave problema de equidad por el que atraviesa actualmente nuestro sistema educativo— tienen quizás su expresión más descarnada en el hecho de que más del 70% de los estudiantes rurales de sexto grado de primaria no es capaz de resolver siquiera todas tareas requeridas del área de Matemática para ubicarse en el nivel previo, lo que supone que ni siquiera podemos afirmar que efectivamente estos estudiantes hayan desarrollado las capacidades establecidas para los grados anteriores en dicha área.

Cuadro 5.5: Sexto grado de primaria: porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño según el área de ubicación de la IE



e.e.: error estándar.

* Diferencia estadísticamente significativa al 5%.

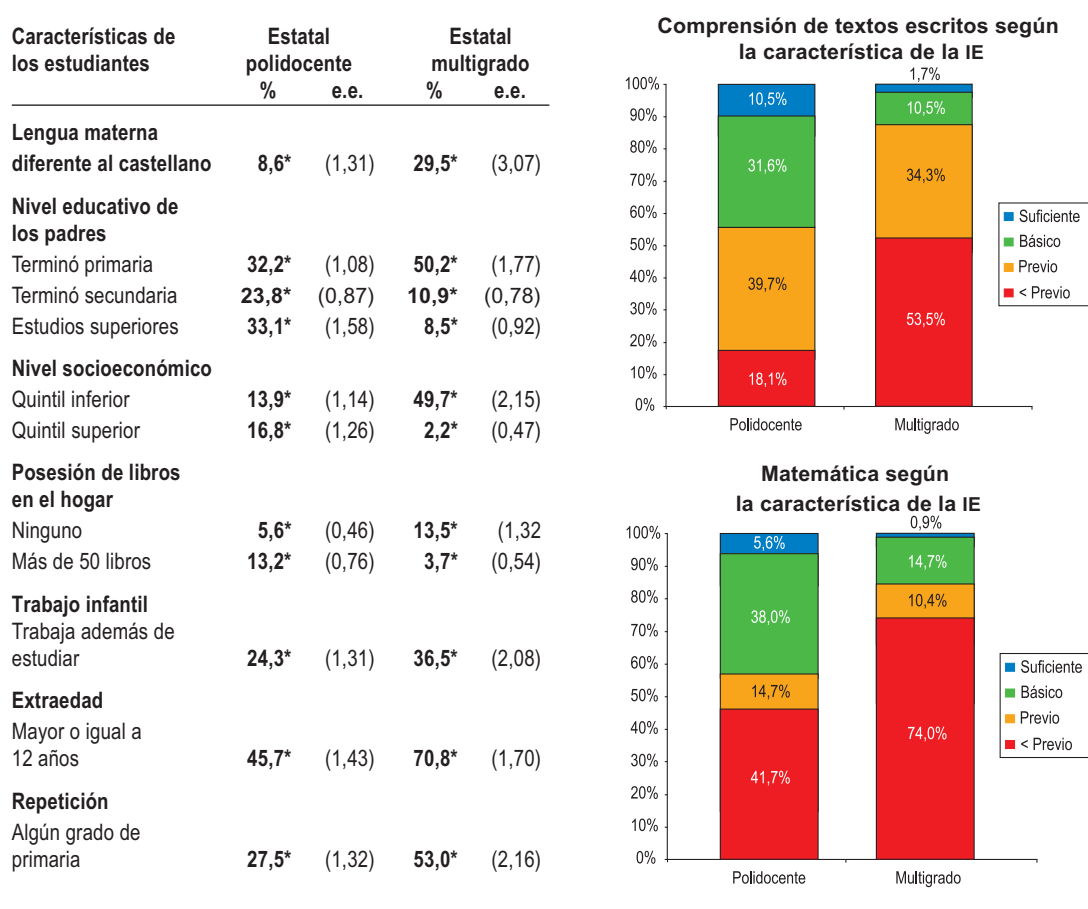
No obstante lo señalado sobre la preocupante situación que presentan los aprendizajes de los estudiantes rurales, es importante llamar la atención sobre los también bajos niveles de logro que presentan sus pares urbanos. En efecto, si observamos la información presentada en el anterior gráfico podemos darnos cuenta de que más de la mitad de los estudiantes urbanos se ubica en el nivel previo o por debajo de este. Es importante no perder de vista esto, en tanto casi siempre se tiende a llamar la atención sobre los pobres resultados educativos de las zonas rurales —lo cual resulta comprensible—, pero se olvida que este problema también está presente en la zona urbana, la que además implica a cerca del 67% de matrícula nacional.

Estos resultados eran de esperarse dada la desigual distribución del ingreso en nuestro país: mientras el 72,5% de la población rural es pobre, este porcentaje disminuye a 43,3% en el caso de la población urbana.⁶⁹ Esto puede observarse si atendemos a algunas variables contextuales de los estudiantes evaluados. Así, se puede apreciar en la información presentada que los estudiantes pertenecientes a las IE de gestión estatal ubicadas en zonas urbanas tienen padres con mejores niveles educativos, pertenecen a estratos socioeconómicos más altos, cuentan con un mayor número de libros en casa, presentan menores índices de repetición y extraedad, y un menor porcentaje de alumnos trabaja además de estudiar. Todas estas variables pueden estar actuando como mediadoras en los efectos de la ruralidad sobre el rendimiento académico.

Mención especial merece la característica sobre la lengua materna diferente del castellano de los estudiantes, aun cuando esta no sea común a la totalidad de estudiantes de las zonas rurales. Como se ha señalado en estudios anteriores (UMC 2004), la lengua materna de los estudiantes es una variable que se asocia significativamente con el rendimiento, aun cuando el efecto de las otras variables estudiadas —incluyendo el nivel socioeconómico— es controlado. Así, los estudiantes cuya lengua materna es el castellano obtienen mejores resultados en comparación con el grupo de estudiantes que tienen una lengua vernácula como lengua materna. Este hecho debe ser comprendido en tanto el programa de Educación Bilingüe Intercultural que atiende escuelas de zonas bilingües de todo el país —buscando que aquellos estudiantes cuya lengua materna es distinta del castellano tengan acceso a una educación de calidad que resulte cultural y lingüísticamente pertinente— todavía no cubre la demanda general de la población vernácula. De otro lado, la política educativa todavía no ha logrado garantizar las condiciones mínimas —asignación de docentes con un manejo adecuado de la lengua de la población a la que enseñan y capacitados para la enseñanza aprendizaje bilingüe, materiales educativos pertinentes, etc.— para que este programa se pueda desarrollar de una manera adecuada.

69. Encuesta Nacional de Hogares 2004. Instituto Nacional de Estadística e Informática.

Cuadro 5.6: Sexto grado de primaria: porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño según la característica de la IE



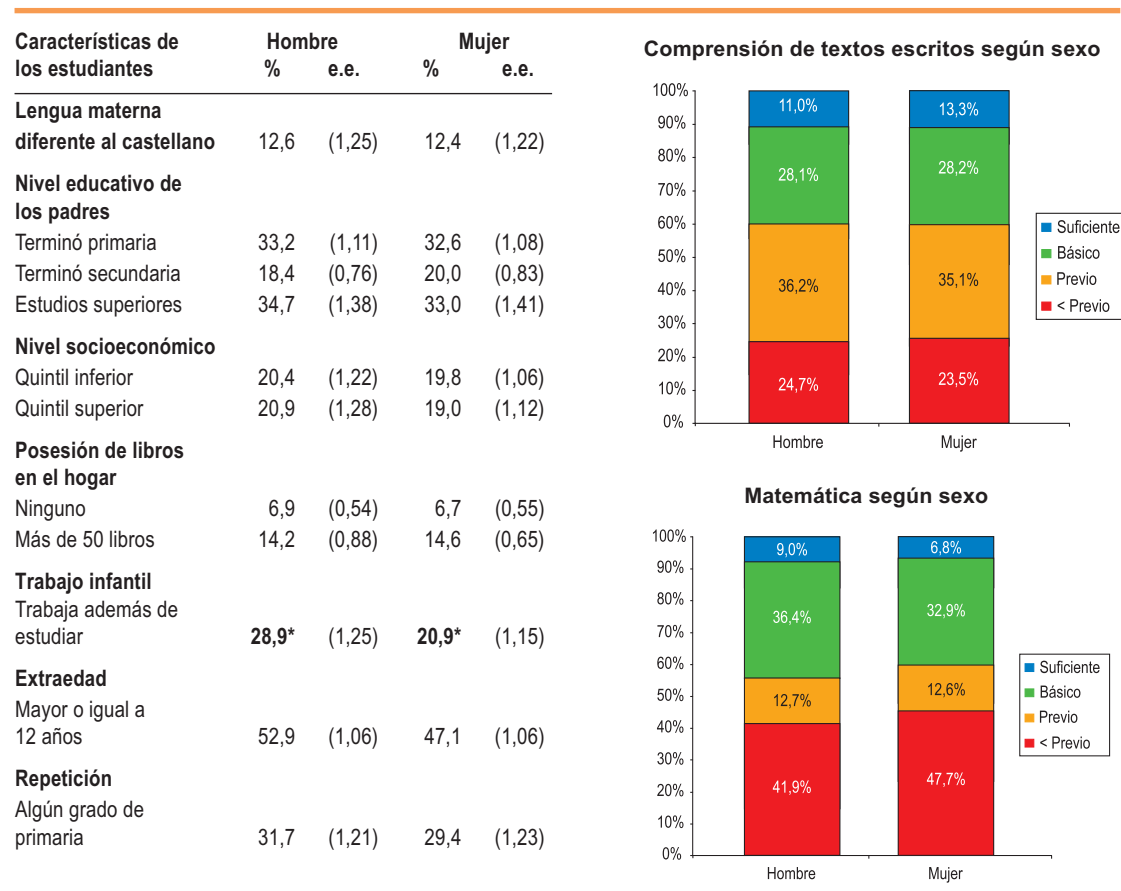
e.e.: error estándar.

* Diferencia estadísticamente significativa al 5%.

Otra división posible al interior de las escuelas estatales, de acuerdo al diseño muestral de la EN 2004, atiende a la característica de la escuela, es decir, si esta es polidocente completa o si es multigrado. De acuerdo al gráfico anterior, cuando comparamos los resultados de los estudiantes que asisten a escuelas de gestión estatal según la característica de la IE, observamos resultados muy similares a los presentados al compararlos según el área de ubicación geográfica. En este caso, los estudiantes de las escuelas polidocentes completas obtienen mejores resultados que sus pares que estudian en escuelas multigrado. Esto no es de extrañar, pues el 89% de las escuelas rurales son multigrados.⁷⁰ Por lo tanto, lo inferido para el estrato urbano y rural es válido también para las IE polidocentes completas y las multigrado.

70. No podemos dejar de señalar que pese a la alta proporción de este tipo de escuelas, solo en los últimos años ha cobrado cierta visibilidad para la política educativa. Así, por ejemplo, recién nos hemos hecho conscientes de que los docentes no han sido formados adecuadamente para enfrentar la enseñanza aprendizaje de uno, dos y hasta seis grados simultáneamente.

Cuadro 5.7: Sexto grado de primaria: porcentaje de estudiantes ubicados en cada uno de los niveles de desempeño según sexo



e.e.: error estándar.

* Diferencia estadísticamente significativa al 5%.

Por último es importante atender al rendimiento de los estudiantes en términos de las diferencias por sexo. Como se puede apreciar en el cuadro anterior, en sexto grado de primaria existen diferencias significativas entre los porcentajes de varones y mujeres que se ubican en el nivel suficiente para ambas áreas evaluadas. En el caso de la prueba de Comprensión de textos escritos, el rendimiento de las mujeres suele ser mejor que el de los varones, mientras que en la prueba de Matemática, los resultados muestran una tendencia opuesta.

Si bien esta tendencia ha sido encontrada también en otros estudios (Benavides 2002; OECD 2002; Espinosa y Torreblanca 2004), es importante tener en cuenta que, tal como señala el informe PISA ⁷¹ 2003, la amplia variación de las diferencias por sexos entre los diferentes países sugiere que las diferencias actuales no son inevitablemente producto de las diferencias entre los chicos y las chicas, y que las políticas y las prácticas eficaces pueden superar lo que durante mucho tiempo se ha considerado como consecuencias ineludibles de las diferencias entre hombres y mujeres respecto a intereses, estilos de aprendizaje e, incluso, capacidades subyacentes (OECD 2004:99).

71. Programa para la Evaluación Internacional de Alumnos (PISA por sus siglas en inglés) de la Organización para la Cooperación y el Desarrollo Económico (OECD).

En esta perspectiva, de acuerdo con las características de los estudiantes según sexo presentadas en el cuadro 5.7, solo aparecen diferencias estadísticamente significativas en la proporción de estudiantes que trabaja además de estudiar, a favor de las mujeres. Esta evidencia abona a lo señalado por PISA, que son los procesos pedagógicos en las aulas y los procesos institucionales de las escuelas son los que estarían influenciando de manera preponderante en las diferencias del rendimiento de hombres y mujeres. Así, mejorar el nivel de participación de los varones en las actividades lectoras, y estimular el interés y el concepto en sí mismo entre las mujeres en Matemáticas, deben ser objetivos de política pedagógica muy serios, si se quiere alcanzar una mayor equidad de género en los resultados educativos (OECD 2003).

5.2.3. RESULTADOS DE LOS ESTUDIANTES EN CONDICIÓN DE POBREZA: RELATIVIZANDO EL EFECTO SOCIOECONÓMICO

Si bien los resultados de los modelos de factores asociados de las evaluaciones anteriores⁷² muestran que gran parte de las diferencias en el rendimiento de los estudiantes de nuestro país⁷³ se explica en una mayor proporción por las características individuales y familiares de los estudiantes mismos antes que por las características de las escuelas a las que asisten. Los análisis de factores asociados también muestran que variables como la composición socioeconómica del alumnado no son suficientes para explicar los diferentes resultados que obtienen las escuelas. Así, por un lado, es posible identificar diferencias significativas en los resultados que alcanzan escuelas que pertenecen a un mismo entorno socioeconómico y, por otro lado, escuelas de diferentes entornos socioeconómicos pueden mostrar resultados similares.

La composición socioeconómica del alumnado no agota la explicación de las diferencias entre las escuelas; algunas variables sobre procesos escolares influyen también en los resultados, lo que evidencia la posibilidad que tiene la escuela de cambiar su entorno y, en algunos casos, hasta de minimizar el impacto negativo de las condiciones socioeconómicas en el aprendizaje de las poblaciones más desfavorecidas a las que atiende.

Este hecho se hace evidente, por ejemplo, al constatar la variabilidad del rendimiento que existe dentro de estos estratos desfavorecidos en la EN 2004. Es decir, a pesar de que los resultados de los más pobres, como se ha podido observar, son bastante bajos, es posible encontrar diferencias al interior de estos. Así, por ejemplo, en el siguiente cuadro podemos observar algunas variables escolares que diferencian a los estudiantes de mayor y menor rendimiento dentro de estos estratos.⁷⁴

72. Los análisis de factores asociados de las evaluaciones anteriores se pueden consultar en: Asmad, et al (2004), Caro, et al (2004), UMC (2004), Benavides (2002).

73. Cabe señalar que esto no es privativo de nuestro país, en términos generales se puede decir que los países de la región tienden a presentar también estos resultados.

74. Para responder esta pregunta, se clasificó a los hogares como pobres y no pobres mediante un modelo de regresión logística en base a la información de la tenencia de algunos bienes en el hogar y el nivel educativo de los padres. Los parámetros de este modelo fueron estimados con la información de la Encuesta Nacionales de Hogares 2003 (ENAH0) la cual permite definir la condición de pobreza de los hogares en base a una línea de pobreza, encontrándose que la tasa de acierto entre los niveles de pobreza estimados con el modelo logístico y con la línea de pobreza era de 76%. Estos parámetros estimados fueron aplicados a la información de la EN 2004 para la clasificación de los hogares de los estudiantes que participaron en dicha evaluación. Una vez identificado el grupo de estudiantes en condi-

Cuadro 5.8: Algunas características de los estudiantes de nivel socioeconómico bajo, según el tercil de rendimiento académico ⁷⁵

Variable estudiada	Comprensión de Textos				Matemática			
	Tercil inferior		Tercil superior		Tercil inferior		Tercil superior	
	Promedio E.E.		Promedio E.E.		Promedio E.E.		Promedio E.E.	
Comunicación con la familia	267,6*	5,4	289,7*	4,1	267,0*	5,5	290,4*	4,2
Relación profesor – alumno	242,7*	5,7	306,2*	3,9	247,7*	5,0	303,4*	4,0
Autoconcepto del estudiante	231,3*	11,5	280,2*	9,2	235,1*	11,3	276,7*	8,7
Habilidades de los docentes	266,8*	6,6	291,7*	5,3	266,0*	8,0	287,0*	5,3
Oportunidades de aprendizaje	90,0*	1,3	92,7*	0,7	76,0*	1,7	86,2*	1,1
Trabajo con pares	291,6	8,7	294,2	6,8	288,0	9,4	286,4	6,8
Satisfacción docente	284,2	7,5	289,2	6,7	279,7	7,6	289,6	6,5
Expectativa docente	273,3	8,5	285,7	8,4	268,0	8,8	290,3	8,9
Manejo de aula	284,0	8,0	285,9	6,8	277,3	7,4	288,3	7,1
Estrategias instruccionales	285,6	8,2	282,3	7,1	287,8	7,7	288,6	7,7
Compromiso con el estudiante	288,9	7,5	294,6	7,7	289,5	7,7	294,4	8,2

* diferencia estadísticamente significativa al 5%.

Como se aprecia en el cuadro 5.8, tanto para el caso de Comprensión de textos escritos como para Matemática, aparecen diferencias estadísticamente significativas en las siguientes variables: Comunicación con la familia, Relación profesor–alumno, Autoconcepto⁷⁶ del estudiante, Habilidades lectoras y/o matemáticas de los docentes y Oportunidades de aprendizaje. Para todas estas variables, tanto en Comprensión de Textos escritos como en Matemática, los mayores promedios aparecen en el grupo de estudiantes pertenecientes al tercil superior de rendimiento.

En el caso de la variable Comunicación con la familia, los estudiantes que tienen una mayor interacción con los miembros de su familia poseen una mejor red de soporte social. Esta probablemente les permite enfrentar de modo más eficiente las dificultades con las cuales se encuentran en la vida cotidiana, y dedicarse de modo más efectivo a las actividades escolares. Además podemos postular que los alumnos que se comunican más con sus padres, en general reciben más apoyo de los mismos para las actividades escolares, lo cual, según algunas investigaciones, impacta positivamente en el rendimiento académico (Fantuzzo, Davis, y Ginsburg, 1995).

Otra variable que sirve para diferenciar a los grupos del tercil superior y del inferior es la relación profesor–alumno. En este caso, se evaluó la percepción que tiene el alumno respecto de las interacciones que realiza con su docente, generando un clima positivo y

ciones de pobreza se les diferenció en función a su rendimiento académico. Esto nos permitió comparar los grupos ubicados en el tercil superior e inferior de rendimiento sobre la base de un conjunto de variables recogidas en los instrumentos de factores asociados.

75. Las diferencias se trabajaron con un nivel de confianza del 95%.

76. Se refiere a las imágenes, sentimientos y rasgos que, en el aspecto académico, tienen los estudiantes de sexto grado sobre sí mismos.

de confianza en el salón de clases. Se indaga por la preocupación del docente por sus alumnos, sobre si fomenta la intervención de los estudiantes por medio de preguntas, y sobre si su trato es amable. Los resultados indican que dentro del grupo de alumnos en condiciones de pobreza, aquellos que pertenecen al tercil superior de rendimiento académico perciben un clima de interacciones más positivas con sus docentes. Estos resultados son congruentes con los hallados en diversas investigaciones donde se encuentra una relación positiva entre el clima de aula y el rendimiento académico de los estudiantes (Chiew y Fraser, 2000; Waxman, Huang, Anderson y Weinstein, 1997). En el caso de los estudiantes más pobres, esto resulta aún más relevante, pues permite compensar algunas carencias debidas a la propia situación socioeconómica.

Con respecto a las diferencias en autoconcepto, diversas investigaciones señalan a esta variable como un elemento decisivo en el proceso de enseñanza aprendizaje al concluir que el éxito o fracaso académico es influenciado no solo por las capacidades del sujeto sino también por sus percepciones como estudiante de Matemática o Comunicación, y por aquello de lo que se siente capaz de hacer en dichas áreas (Bouchey y Harter, 2005; Saura, 1995). Con esto no queremos afirmar de ninguna manera que el autoconcepto causa el rendimiento académico, solo señalamos que ambas variables se encuentran relacionadas. Probablemente, como lo indica Marsh (1990), las relaciones entre ambas variables son complejas y hay una interdependencia entre ambas. El autoconcepto académico se desarrolla y mejora con las experiencias de progreso y éxito académico. Es decir, el estudiante, al constatar empíricamente que es bueno en Matemática y/o en Comunicación, se conceptualiza a sí mismo como bueno en esas áreas. Por otro lado, el estudiante con un buen autoconcepto en dichas áreas académicas, se esforzará e interesará más en ellas, lo cual impactará positivamente en su rendimiento académico.

Como observamos en la variable habilidades de los docentes, los estudiantes con mejor rendimiento, en general, tienen docentes con mejores resultados en los instrumentos de factores asociados que intentaron aproximarse a sus habilidades en las áreas de Comunicación y Matemática. Si asumimos que resulta difícil enseñar lo que no se domina, los docentes con mejor dominio del área estarían en mejores condiciones para fomentar el desarrollo de las capacidades en sus alumnos. Por lo tanto, es más probable que un docente competente logre generar adecuados aprendizajes en sus alumnos, que uno que no domina los contenidos del área. Al respecto, Fillmore y Snow (2000) señalan que un docente de Comunicación que domina su materia, además de tener un buen conocimiento del desarrollo del lenguaje, tiene mayores posibilidades de lograr estudiantes competentes en esta área. Por su parte, Hill, Rowan y Loewenberg (2005) encuentran en sus investigaciones resultados similares en el área de Matemática.

Finalmente, en el caso de las Oportunidades de Aprendizaje,⁷⁷ dentro del grupo de pobreza, los alumnos con un mejor rendimiento han estado expuestos a un mayor número de contenidos curriculares. Es decir, estos estudiantes han tenido mayores oportunidades de aprendizaje, lo que ha impactado positivamente en su rendimiento académico. Al respecto, los trabajos de Cueto, Ramírez, León y Pain (2003), encuentran también asociaciones positivas entre las oportunidades de aprendizaje y el rendimiento académico.

Como señalamos al inicio de este apartado, la composición socioeconómica del alumnado no es la única variable que explica sus diferencias en el rendimiento académico. Las

77. Este indicador se refiere al porcentaje de las competencias desarrolladas en el aula, tomando como referencia el total de competencias evaluadas en las pruebas aplicadas en la EN 2004.

variables sobre procesos escolares aquí estudiadas influyen en los resultados, lo que evidencia la posibilidad que tiene la escuela de cambiar su entorno y así crear mejores condiciones que favorezcan los aprendizajes. En algunos casos, hasta sería posible minimizar el impacto de las condiciones socioeconómicas sobre el aprendizaje de las poblaciones más desfavorecidas a las que atiende.

En este sentido, es importante que la escuela tome en cuenta estos factores y que propicie mejores condiciones en el entorno de los estudiantes, por ejemplo, fomentando la comunicación e interacción positiva con los padres, y su apoyo e interés en el desempeño escolar de sus hijos. Asimismo, el asegurar un clima positivo en el aula, donde prime la cooperación y colaboración, puede generar mejores aprendizajes. Otra manera de contrarrestar el factor de riesgo que conlleva la condición de pobreza, implica asegurar que estos estudiantes tengan docentes que realmente dominen los contenidos, enfoque y didácticas de las áreas, de tal manera que sean capaces de fomentar aprendizajes en los estudiantes. Este hecho es bastante positivo, pues si los estudiantes tienen mayores y mejores oportunidades de aprendizaje, se podría, parcialmente, contrarrestar los impactos negativos que tiene la condición de pobreza sobre su rendimiento. El generar todas estas condiciones que contribuyan de manera positiva al desarrollo de las capacidades de los estudiantes en ambas áreas a su vez contribuiría a formar un autoconcepto académico positivo, que reforzará el interés del alumno por su propio aprendizaje.



En el marco de la EN 2004, se realizó un estudio con el fin de identificar variaciones del rendimiento académico de los estudiantes de sexto grado de primaria y quinto grado de secundaria en las áreas de Matemática y Comunicación a lo largo del tiempo. Se eligieron esos grados en tanto son grados finales de cada nivel y nos proporcionan información valiosa sobre los resultados obtenidos por los estudiantes al culminar cada una de las etapas de la Educación Regular Básica.

Con el fin de evaluar la presencia de cambios en el rendimiento de dichas poblaciones, se utilizó la información proporcionada por las pruebas de rendimiento aplicadas en el marco de las evaluaciones nacionales de los años 1998 y 2004, cuyas características se resumen en el cuadro que se presenta a continuación.

Cuadro 6.1: Características de las evaluaciones nacionales usadas en las comparaciones

Evaluación	Modelo	Áreas	Ámbito
CRECER 1998	Normas	Matemática	Urbano Polidocentes completas
EN 2004	Criterios	Comunicación	Nacional

Como se puede apreciar, la evaluación CRECER 1998 se efectuó bajo un modelo de normas.⁷⁸ Sus resultados son representativos de las instituciones educativas urbanas, que además se caracterizan por ser polidocentes completas. Por su parte la Evaluación Nacional 2004 empleó un modelo de criterios y tiene representatividad a nivel nacional.

El estudio de comparación de rendimientos está dado en el marco del modelo de normas utilizado en el año 1998.⁷⁹ Se aplicó en sexto de primaria la misma prueba (cuadernillo CT98, compuesto por 32 ítemes) de Comunicación Integral y de Lógico Matemática a una submuestra de estudiantes evaluados en el año 2004. Para quinto de secundaria, en CRECER 98 se utilizaron dos cuadernillos de 25 ítemes cada uno en matemáticas. En el año 2004 se combinaron ambos cuadernillos en una sola prueba conformada por 25 ítemes, y fue aplicada a una submuestra de estudiantes. En el caso de Comunicación, se aplicó a una submuestra de estudiantes del 2004 la misma prueba utilizada en CRECER 98, conformada por 32 ítemes.

78. El modelo de normas busca mostrar la posición relativa entre diferentes grupos, pero no ofrece información sobre el nivel de desempeño de los mismos.

79. Al ser este modelo bajo el cual se analizaron las pruebas de CRECER 98 se decidió mantenerlo para realizar la comparación de los cuadernillos compartidos.

6.1. Metodología

Para garantizar la comparabilidad de los resultados entre las diversas evaluaciones se realizaron tres pasos previos: la selección de muestras comparables y ponderación, el análisis del funcionamiento diferencial de los ítemes, y el escalamiento de las pruebas.

6.1.1. SELECCIÓN DE MUESTRAS COMPARABLES Y PONDERACIÓN

La evaluación CRECER 98 se aplicó a una muestra de estudiantes—representativa de cuarto y sexto de primaria y de cuarto y quinto de secundaria— de instituciones educativas polidocentes urbanas a nivel nacional. En total, se aplicaron los instrumentos en 576 IE de primaria y 566 IE de secundaria, con un esperado de 17 370 estudiantes en cada grado evaluado.

Dadas las diferencias en los tamaños de la muestra CRECER 98 y la submuestra de la EN 2004, se procedió a generar muestras equivalentes de manera que puedan ser comparables.

Considerando que la principal diferencia entre la población objetivo de ambas poblaciones se refiere a la restricción de la aplicación de CRECER 98 a solo IE polidocentes completas del ámbito urbano, se trabajó con ponderaciones de tal manera que reflejaran la distribución real del alumnado en los estratos dados por el cruce de las variables gestión (estatal / no estatal) y ubicación de la IE (Lima Metropolitana/ Costa Central —sin Lima— / Costa Norte / Costa Sur /Sierra/Selva). Con ello se garantizó la equivalencia en la composición de la submuestra de la EN 2004 con la muestra de la evaluación CRECER 98.

6.1.2. ANÁLISIS DEL FUNCIONAMIENTO DIFERENCIAL DE LOS ÍTEMES (FDI)

Uno de los requisitos para realizar la comparación del rendimiento entre dos o más grupos es asegurar la invarianza de la dificultad de los ítemes de la prueba a lo largo del tiempo. Esto supone que el grado de dificultad de cada ítem debe de mantenerse constante,⁸⁰ independientemente del grupo poblacional en el que se aplique. En caso de no cumplirse con este supuesto, se dice que el ítem presenta un funcionamiento diferencial. Para realizar dicho análisis se revisó el comportamiento de los ítemes a lo largo del tiempo y fueron elegidos solo aquellos que cumplían con este supuesto.

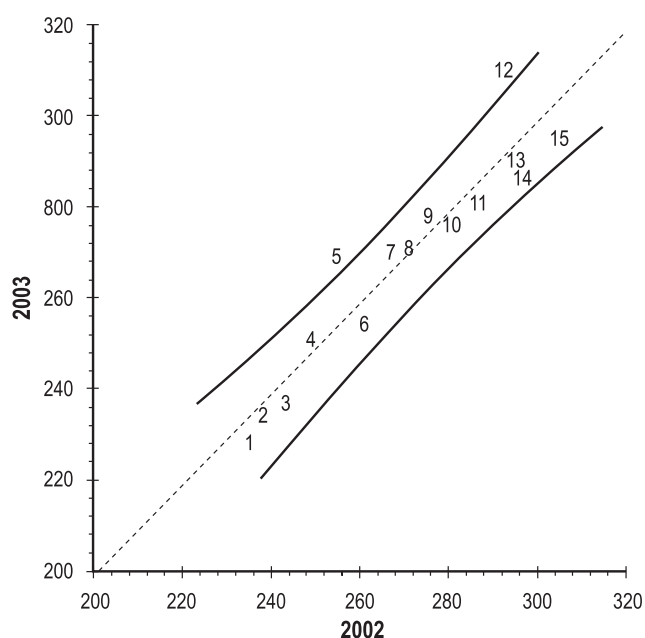
Se utilizó el método gráfico basado en el diagrama de dispersión para determinar qué ítemes se emplearían para poner las pruebas en la misma escala. Este método consiste en obtener de forma independiente las dificultades de los ítemes comunes en cada una de las muestras en las que fueron aplicados. Luego se llevan estos ítemes a un plano de coordenadas que muestra en el eje x las dificultades obtenidas en la población 1. En el eje y aparecen las dificultades calculadas para la población 2. Al forzar una regresión con una pendiente de uno, se puede ver qué tanto se alejan los ítemes de esa línea de regresión y establecer cuáles son los que mejor se ajustan al supuesto de invarianza. Si los ítemes se alejan mucho de la línea de ajuste, están funcionando de forma distinta en

80. La dificultad del ítem debe mantenerse dentro de ciertos márgenes de error al comparar su dificultad en dos poblaciones diferentes. En nuestro caso, se refiere a los dos momentos en los cuales se aplicó la prueba.

ambas poblaciones (Hambleton, Swaminathan y Rogers, 1991; Ingebo, 1997). En esos casos, el ítem no fue incluido en el proceso de poner las pruebas aplicadas a ambas poblaciones en la misma escala.

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.⁸¹ Si un ítem cae fuera de dichas líneas de control, estaremos seguros de que no cumple con los requerimientos de invarianza.

Por ejemplo, supongamos que se ha aplicado una prueba de aritmética compuesta por quince ítems a dos muestras representativas de alumnos de tercer grado de primaria. Una de las muestras fue tomada en el año 2002 y la otra en el 2003. Se desea comparar el rendimiento de la población de alumnos en ambos años. Para ello se debe determinar qué ítems se mantienen invariantes en ambos años. Se calcula la dificultad de los ítems para cada una de las muestras por separado y se construye con esos datos un diagrama de dispersión:



Como se observa en el gráfico anterior, dos ítems caen fuera de las líneas de control, el 5 y el 12. Ambos ítems no se considerarían dentro del proceso de poner dos pruebas aplicadas a dos poblaciones distintas en la misma escala. Nótese 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.

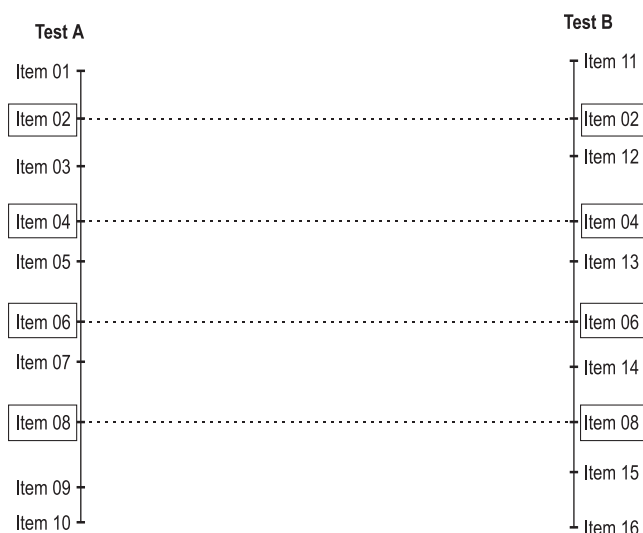
81. Cuando se estima la dificultad de un ítem, en realidad solo nos aproximamos a su valor. No lo podemos conocer exactamente. Siempre tiene asociado cierto error de estimación (error estándar). En función de este error estándar se pueden trazar intervalos dentro de los cuales es probable que se encuentre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

6.1.3. ESCALAMIENTO DE LAS PRUEBAS

En el marco de la Teoría de Respuesta al Ítem (TRI), al ser las puntuaciones invariantes con respecto a las personas y el instrumento, no se necesita equiparar las puntuaciones, sino simplemente ponerlas en la misma escala. Recuérdese que las puntuaciones del rasgo latente tienen un origen y escala de medición arbitraria. 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. Esta conversión es análoga a pasar de centímetros a pulgadas. Para ello existen diversos procedimientos utilizados en los Modelos Rasch y de Teoría de Respuesta al Ítem, a fin de poner las pruebas aplicadas a poblaciones diferentes en la misma escala.⁸² Uno de los métodos más utilizados consiste en el uso de ítems comunes, llamado también método de los ítems de anclaje. Este fue el método que se utilizó para poner las pruebas en la misma escala. A continuación, se describe la lógica que sustenta la validez de este método.

Supongamos que se han construido dos pruebas (Test A y B), cada una de diez ítems:



Si se suman los ítems de ambos tests, podemos observar que en total se tienen 16 ítems distintos. Esto se debe a que hay cuatro ítems comunes a ambos tests (Ítems 02, 04, 06 y 08) los cuales se utilizarán para poner ambos instrumentos en la misma escala de medición. Esto se efectúa mediante el cálculo de una constante de escalamiento, la cual es más confiable mientras más ítems en común existan entre las dos pruebas (Ingebo, 1997). Además es posible aplicar el procedimiento de escalamiento concurrente o conjunto. Este consisten en construir una matriz con las respuestas de los alumnos evaluados en los dos momentos y analizarlas de manera conjunta, como si se tratara de una única matriz.

82. Por ejemplo, los métodos basados en la media y la desviación estándar, los basados en la curva característica de los ítems, los de las dificultades fijas y de calibración concurrente (Navas, 1996; Kolen y Brennan, 2004).

La literatura no especifica exactamente cuántos ítemes en común debe haber. Lo usual es señalar que se necesitan entre 20 o 30 ítemes en común, o tener por lo menos un 20% de ítemes de anclaje (Kolen y Brennan, 2004). Por su parte, algunos autores como Reckase (1979; en Navas, 1996) señalan que al trabajar con el modelo Rasch y muestras superiores a 300 personas, 15 ítemes en común pueden ser suficientes.

La efectividad de la constante de escalamiento depende además de lo similares que sean los ítemes comunes a los demás ítemes que conforman la prueba. En otras palabras, los ítemes comunes deben ser una muestra representativa del conjunto de ítemes de la prueba en su contenido y cubrir un amplio rango de dificultad.

La manera como se colocó en la misma escala las pruebas aplicadas en el marco de las diferentes evaluaciones nacionales implicó identificar aquellos ítemes comunes que no mostraban un marcado funcionamiento diferencial y utilizarlos como ítemes de anclaje en el programa informático Winsteps versión 3.57.2, que utiliza la información proveniente de los ítemes de anclaje para poner las medidas de dificultad de los ítemes y habilidad de las personas en una escala común. (Linacre, 2005).

6.2. Resultados

6.2.1. SEXTO DE PRIMARIA

Cuadro 6.4: Sexto de primaria: resultados generales

Área	1998		2004		Diferencia 99%
	media	error	media	error	
Comunicación Integral	308,5	1,54	303,2	2,71	—
Lógico Matemática	306,3	1,60	311,9	3,11	—

Como observamos en el cuadro anterior, no hay diferencias estadísticamente significativas entre las medias aritméticas de los alumnos de sexto de primaria que rindieron las pruebas en 1998 y los que las rindieron en 2004. Esto nos estaría indicando que el rendimiento promedio de los estudiantes de IE urbanas polidocentes completas no ha variando entre los años 1998 y 2004 según los contenidos evaluados en CRECER 98, tanto para Comunicación Integral como para Lógico Matemática.

A fin de investigar con mayor precisión la existencia de diferencias en el rendimiento entre ambos años, se comparó el rendimiento a nivel de algunos grupos poblacionales, lo que implicó comparar diferencias dentro de las IE estatales, las no estatales, las de Lima y, finalmente, las de provincias.

Cuadro 6.5: Comunicación Integral: resultados por grupos comparados

Grupo	1998		2004		Diferencia 99%
	media	error	media	error	
IE estatales	298,8	1,61	291,0	2,89	—
IE no estatales	357,5	4,26	358,2	5,73	—
Lima	330,4	2,70	323,4	4,98	—
Provincias	295,9	1,89	292,2	3,06	—

El cuadro 6.5 muestra que no hay diferencias estadísticamente significativas entre las medias aritméticas en Comunicación Integral de los estudiantes de sexto de primaria que rindieron la prueba en el año 1998 y los que la rindieron en el año 2004 en ninguno de los grupos comparados.

Cuadro 6.6: Lógico Matemática: resultados por grupos comparados

Grupo	1998		2004		Diferencia 99%
	media	error	media	error	
IE estatales	296,6	1,72	300,3	3,45	—
IE no estatales	354,6	4,12	362,3	6,05	—
Lima	323,3	2,75	334,3	5,15	—
Provincias	296,3	1,98	300,3	3,67	—

De igual manera, como se puede apreciar en el cuadro 6.6, no existen diferencias estadísticamente significativas al comparar los promedios de los alumnos de sexto de primaria que realizaron la prueba de Lógico Matemática en el año 1998 y los que la realizaron en el año 2004 en ninguno de los grupos estudiados.

Estos resultados en Comunicación Integral y Lógico Matemática nos estaría indicando que no ha habido una mejora significativa en el rendimiento promedio de los alumnos de IE urbanas polidocentes completas en ninguno de los subgrupos contrastados entre los años 1998 y 2004, según los contenidos evaluados en CRECER 98.

6.2.2. QUINTO DE SECUNDARIA

Cuadro 6.7: Quinto de secundaria: resultados generales

Área	1998		2004		Diferencia 99%
	media	error	media	error	
Comunicación	504,4	1,06	501,8	1,79	—
Matemática	507,2	1,34	497,3	3,04	↓

Cuando se comparan los resultados en las pruebas de rendimiento aplicadas a los estudiantes de quinto de secundaria, se aprecia en el cuadro 6.7 que no hay diferencias significativas en el rendimiento en Comunicación. En cambio, se observa una disminución significativa en el rendimiento promedio en Matemática al comparar a los estudiantes evaluados en el año 2004 con los evaluados en el año 1998.

Cuadro 6.8: Comunicación: resultados por grupos comparados

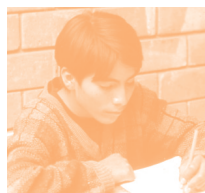
Grupo	1998		2004		Diferencia 99%
	media	error	media	error	
IE estatales	494,3	1,13	492,2	2,13	—
IE no estatales	541,1	2,52	530,9	3,18	—
Lima	519,1	1,82	513,3	1,48	—
Provincias	495,4	1,30	496,4	2,15	—

El cuadro 6.8 muestra que no existen diferencias estadísticamente significativas al comparar los promedios de los estudiantes de quinto de secundaria que rindieron la prueba de Comunicación en el año 1998 y los que la realizaron en el año 2004 en ninguno de los grupos estudiados.

Cuadro 6.9: Matemática: resultados por grupos comparados

Grupo	1998		2004		Diferencia 99%
	media	error	media	error	
IE estatales	494,9	1,40	482,5	3,44	↓
IE no estatales	549,1	3,21	547,4	6,34	—
Lima	519,5	2,36	510,7	6,59	—
Provincias	499,4	1,62	491,1	3,43	—

La diferencia significativa que se observó en el rendimiento promedio en Matemática aparece solo en las IE estatales, las cuales muestran una disminución de su rendimiento promedio en el año 2004 frente al año 1998. El resto de grupos de estudiantes comparados no muestra ninguna diferencia estadísticamente significativa. Esto quiere decir que las diferencias en el rendimiento promedio observadas en el cuadro 6.7 se explicarían por la disminución en el rendimiento de los estudiantes de las instituciones estatales.



ANEXOS

Anexo 1: Diseño muestral de la EN 2004*

Introducción

En este anexo se presenta el diseño muestral (población objetivo, marco muestral, tamaño de muestra, procedimiento de selección, cálculo de los errores estándar y factores de expansión) utilizado en la Evaluación Nacional 2004 (EN 2004).

El objetivo de la evaluación es medir los rendimientos escolares logrados por los estudiantes de segundo y sexto grados de primaria y de tercer y quinto grados de secundaria a nivel nacional. Esta medición se realiza por medio de la aplicación de pruebas especialmente diseñadas por el equipo de Evaluación de la UMC a una muestra de estudiantes que representa a la población matriculada y que asiste a alguna institución educativa (IE) de educación primaria o secundaria bajo la modalidad de escolarizada de menores.⁸³ Estas pruebas fueron aplicadas entre el 15 y el 19 de noviembre de 2004.

1. Población objetivo

La población objetivo se puede dividir en cuatro grupos según el grado escolar de los estudiantes evaluados, todos pertenecientes a IE estatales o no estatales en la modalidad escolarizada de menores a escala nacional. Las poblaciones de estudio estaban conformadas por:

1. Estudiantes de segundo grado de primaria
2. Estudiantes de sexto grado de primaria
3. Estudiantes de tercer grado de secundaria
4. Estudiantes de quinto grado de secundaria

Además de considerar al estudiante como unidad de análisis, se consideraron otras poblaciones con el fin de recolectar información sobre variables asociadas con el rendimiento del mismo. El estudio de estas poblaciones y su relación con el rendimiento de los estudiantes permite identificar los factores asociados al rendimiento. En tal sentido, se aplicaron una serie de instrumentos adicionales:

83. Modalidad escolarizada es la modalidad de asistencia regular al centro educativo durante el año escolar que, entre otras cosas, exige un mínimo de asistencia para no ser eliminado. En oposición a esta modalidad, la no escolarizada es semejante a un programa no presencial que no requiere asistencia regular a lo largo del año escolar. La modalidad de menores se refiere a que está dirigida a la población de quince años o menos. La población escolar para esta modalidad se estima en 778 598; 611 923; 378 725; y 296 572 estudiantes de segundo y sexto grado de primaria y de tercer y quinto de secundaria en el año 2003, respectivamente. Para mayor detalle de la distribución de este universo, véanse Cuadros 1.1 al 1.4.

* Participó en la redacción de este anexo Daniela Zacharías Zanotti

- Cuestionarios a docentes de los estudiantes evaluados en las áreas de Lógico Matemática/Matemática y Comunicación Integral/Comunicación. Los docentes de sexto grado de primaria y de quinto grado de secundaria del área de Personal Social / Desarrollo Social⁸⁴ fueron enfrentados solamente al cuestionario de concepciones pedagógicas.
- Una guía del aula y del centro educativo con variables acerca de la escuela (infraestructura y recursos) y del alumnado que la compone, asistencia a clase, cantidad de estudiantes por grado, entre otros.
- Además, los estudiantes que rindieron la prueba en sexto grado de primaria y en quinto grado de secundaria respondieron un cuestionario que recogía información acerca de aspectos personales, como el nivel socioeconómico, la relación con sus docentes y compañeros, las estrategias de aprendizaje, entre otros.

2. Marco muestral

Para la selección de la muestra de IE se utilizó la base de datos del Padrón de Centros Educativos 2003, elaborado por la Unidad de Estadística Educativa del Ministerio de Educación del Perú.

Una semana antes de la aplicación en campo, se llevó a cabo la validación de los datos de este padrón mediante el envío de formatos de verificación de datos a cada una de las Unidades de Gestión Educativa Local (UGEL) correspondientes a la muestra de IE seleccionadas en los que se solicitaba la matrícula de cada uno de los grados, además del número de secciones por cada uno de ellos. La información fue recibida en aproximadamente el 81% de IE seleccionadas.

3. Método de muestreo

En el diseño muestral se empleó una muestra probabilística⁸⁵, poli-etápica, de conglomerados⁸⁶ y estratificada⁸⁷. A continuación, se detalla el proceso de selección de la muestra bajo los criterios mencionados:

84. Áreas desde las cuales se ha investigado el eje de Formación Ciudadana, en primaria y secundaria respectivamente.

85. En una muestra probabilística cada unidad del universo de estudio tiene una probabilidad de selección conocida y superior a cero. Este tipo de muestra permite calcular los errores muestrales y así poder generalizar los resultados de la muestra a la población de estudio, especificando la precisión de estas inferencias.

86. En términos de muestreo, a las instituciones educativas se los denomina «conglomerados» ya que conforman un conjunto que agrupa a las unidades de análisis (estudiantes).

87. El muestreo estratificado asegura una mayor confiabilidad de la muestra, pues disminuye la varianza de las estimaciones. Una estratificación es más eficiente cuanto más homogéneas sean las unidades que pertenecen a un mismo estrato y más heterogéneos los estratos entre sí. Aun sin cumplir estrictamente con estas características, cualquier estratificación mejora la calidad de las estimaciones, con la única condición de que toda unidad de muestreo pertenezca a un estrato y solo a uno, y la unión de todos los estratos conformen la población total de estudio. La estratificación realizada fue de dos tipos: estratificación explícita, que consta de estratos con fines representativos —de los que se espera obtener estimaciones confiables para comparaciones entre ellos mismos o seguimiento—, y estratificación implícita, en la que los estratos se utilizan con fines de dispersión, lo que mejora la eficiencia del muestreo.

3.1. SELECCIÓN DE IE

La unidad primaria de muestreo es la IE. Esto quiere decir que en una primera etapa se seleccionan IE mediante un muestreo estratificado sistemático con inicio aleatorio y probabilidad proporcional al tamaño (cantidad de estudiantes).⁸⁸

Se definieron estratos combinando criterios que generan grupos de centros educativos homogéneos entre sí. El primer criterio de estratificación fue el tipo de gestión de la IE: estatal o no estatal. Dentro del estrato de las IE de gestión estatal se crearon nuevas divisiones (sub estratos), según el departamento y el área a la que pertenece la IE (urbana o rural) y, en el caso de primaria, se creó una división entre escuelas polidocentes completas o polidocentes multigrado/unidocentes.

A partir del marco muestral, se generó una tabla de selección para cada uno de los estratos dependiendo de las celdas que conforman el diseño complejo, es decir, se elaboró una tabla en la que se especifica el número de estudiantes que se elegirán de cada sub estrato.

Los procedimientos de selección proporcional al tamaño son sistemáticos con arranque aleatorio. Los tamaños de las IE se definen tomando en cuenta un promedio de estudiantes según la característica de la escuela. Así, para las polidocentes completas se seleccionan 30 estudiantes por IE y para las IE unidocentes y polidocentes multigrado, se seleccionan quince estudiantes.

3.2. SELECCIÓN DE SECCIONES DENTRO DE CADA IE SELECCIONADA

Dentro cada una de las escuelas se realizó, como segunda etapa de selección, un muestreo de secciones. Se seleccionaron de manera aleatoria como máximo dos secciones.

3.3. SELECCIÓN DE ESTUDIANTES DENTRO DE CADA CENTRO EDUCATIVO

Como se señaló, se seleccionó un máximo de 30 estudiantes en cada IE. Si la IE tenía dos secciones seleccionadas, se procedía a elegir de manera aleatoria un máximo de 15 estudiantes en cada una de ellas. Si solo se trataba de una sección, se seleccionaba un máximo de 30 estudiantes de la misma. Con el fin de garantizar una selección aleatoria de los estudiantes se capacitó a los examinadores en el uso de la cartilla de muestreo, la cual simula un muestreo sistemático con inicio aleatorio, tomando como marco muestral la lista de asistencia de la IE.

3.3.1. Reemplazo de IE

El reemplazo de IE se realizó en dos momentos: uno después de la verificación y otro, por problemas surgidos en campo, al momento de la aplicación. En el primer caso, la IE se reemplazó por su inmediata anterior o posterior en las listas de selección de IE, es decir, se reemplaza por una IE con igual o similar probabilidad de selección. El segundo caso, el

88. El tamaño de las IE para cada uno de los niveles evaluados fue recategorizado de la siguiente manera: a los centros educativos con 10 estudiantes se les asignó un MOS (*Measure of size*) de 10; a aquellas IE que contaban con más de 10 pero con menos de 21 estudiantes en el grado evaluado se les asignó un MOS de 20. Al resto de IE se les asignó un MOS de 30. Esta medida de tamaño fue la usada para fijar la probabilidad de selección de cada centro educativo. Nótese que según este procedimiento, las IE muy pequeñas obtuvieron alguna probabilidad mayor de ser elegidas.

realizado durante el operativo el reemplazo, además de mantener las exigencias técnicas, dependió de la ubicación de la IE.

4. Tamaño y distribución del universo

En los siguientes cuadros se presenta la distribución del universo de cada grupo de estudio y de la muestra obtenida.

Cuadro 1.1: Distribución del universo de estudiantes de segundo grado de primaria

Departamento	Estatal								No estatal	
	Urbano				Rural					
	Unidocentes/ Multigrado		Polidocentes completos		Unidocentes/ Multigrado		Polidocentes completos			
	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE
Amazonas	1 057	62	4 259	81	9 679	882	755	23	623	49
Ancash	737	59	14 879	244	11 422	1 184	4 753	132	3 224	145
Apurímac	396	23	6 704	113	7 231	609	3 245	89	308	17
Arequipa	690	65	13 691	247	2 077	318	1 119	36	7 795	389
Ayacucho	634	74	10 488	189	9 707	989	2 050	71	930	50
Cajamarca	1 037	47	10 929	188	32 942	2 874	7 649	231	2 356	171
Callao			13 547	116					5 847	342
Cusco	658	58	16 403	214	17 737	1 198	4 015	103	3 283	139
Huancavelica	849	52	4 661	86	10 897	975	1 982	44	223	19
Huánuco	487	20	8 858	125	16 998	1 249	3 597	79	1 170	61
Ica	819	40	10 891	133	1 272	192	640	14	2 929	176
Junín	158	64	16 362	232	11 668	1 235	4 439	103	3 302	220
La Libertad	224	56	22 275	265	13 358	1 107	2 281	70	7 064	314
Lambayeque	032	17	17 787	143	4 616	406	3 340	82	4 019	183
Lima	715	172	103 738	1 026	2 691	394	2 761	72	56 180	2 929
Loreto	275	43	15 104	173	16 355	1 816	2 557	63	1 653	46
Madre de Dios	48	5	1 952	26	716	139	89	3	82	5
Moquegua	106	14	2 190	43	318	95	229	11	367	18
Pasco	298	17	3 964	65	3 619	495	412	16	358	29
Piura	2 122	93	24 989	317	14 890	1 357	3 644	74	5 002	246
Puno	615	38	15 963	226	11 916	1 255	5 244	192	2 322	154
San Martín	675	42	12 280	190	9 382	847	2 047	57	739	64
Tacna	285	16	4 123	61	232	67	246	10	1 115	40
Tumbes	43	5	3 917	63	190	51	188	15	390	20
Ucayali	217	10	9 243	68	6 734	613	1 769	28	245	16
Total	22 177	1 092	369 197	4 634	216 647	20 347	59 051	1 618	111 526	5 842

Cuadro 1.2: Distribución del universo de estudiantes de sexto grado de primaria

Departamento	Estatal								No estatal	
	Urbano				Rural					
	Unidocentes/ Multigrado		Polidocentes completos		Unidocentes / Multigrado		Polidocentes completos			
	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE
Amazonas	831	62	3 653	79	5 213	818	513	23	453	42
Ancash	448	53	13 656	244	5 823	756	3 869	132	2 387	117
Apurímac	361	21	6 558	113	3 752	321	2 667	88	247	13
Arequipa	512	55	13 728	243	1 330	266	982	36	6 423	313
Ayacucho	782	60	8 993	189	4 338	572	1 459	70	618	40
Cajamarca	804	47	9 701	188	19 984	2 789	5 676	231	1 418	147
Callao			12 756	116					4 400	282
Cusco	1 094	46	15 239	212	8 020	724	3 125	103	2 457	117
Huancavelica	605	48	4 041	86	6 403	731	1 621	44	127	14
Huánuco	317	17	8 398	124	7 138	805	2 761	79	866	51
Ica	668	39	10 130	133	1 085	193	523	14	1 889	136
Junín	853	64	15 933	231	7 042	1 190	3 633	103	2 645	174
La Libertad	655	54	18 935	263	6 967	951	1 615	69	5 533	263
Lambayeque	737	17	14 529	144	2 748	400	2 524	82	2 544	153
Lima	3 270	164	102 557	1 021	1 995	370	2 392	72	43 861	2 413
Loreto	810	42	12 448	173	9 668	1 719	1 998	63	1 503	43
Madre de Dios	26	4	1 739	26	473	129	74	3	25	3
Moquegua	88	14	2 060	41	298	82	207	10	262	14
Pasco	162	16	3 854	65	2 456	480	382	16	279	19
Piura	1 516	92	21 029	317	9 252	1 345	2 848	73	4 081	218
Puno	483	35	15 009	226	9 776	1 206	4 716	192	1 906	143
San Martín	493	41	10 673	189	5 277	813	1 497	57	536	61
Tacna	231	16	3 985	61	208	69	233	10	934	32
Tumbes	34	4	3 272	62	178	44	181	17	280	17
Ucayali	121	8	7 627	68	3 556	563	1 193	28	176	14
Total	15 901	1 019	340 503	4 614	122 980	17 336	46 689	1 615	85 850	4 839

Cuadro 1.3: Distribución del universo de estudiantes de tercer grado de secundaria

Departamento	Estatál				No estatal	
	Urbano		Rural			
	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE
Amazonas	3 928	85	1 280	58	417	21
Ancash	9 211	182	3 222	113	2 035	72
Apurímac	5 521	83	617	25	270	11
Arequipa	13 115	158	845	25	6 116	206
Ayacucho	5 782	99	741	35	910	41
Cajamarca	10 041	152	6 387	270	2 134	94
Callao	8 182	50			3 555	157
Cusco	12 656	123	1 583	51	2 423	92
Huancavelica	4 515	100	2 330	94	358	20
Huánuco	7 028	78	1 596	55	897	35
Ica	8 826	74	360	19	1 644	58
Junín	10 145	97	2 267	54	2 899	120
La Libertad	14 168	166	1 838	108	3 877	140
Lambayeque	12424	80	1 342	46	2 703	76
Lima	78 541	602	1 395	48	38 350	1 351
Loreto	7 609	61	937	46	1 287	24
Madre de Dios	952	11	87	6	50	3
Moquegua	1 859	31	292	10	252	11
Pasco	3 119	47	393	20	360	14
Piura	16 718	193	4 160	165	3 384	127
Puno	11 372	95	2 533	78	1 930	75
San Martín	7 566	95	1 023	38	357	12
Tacna	2 898	44	181	11	1 012	24
Tumbes	2 413	28	136	10	288	11
Ucayali	5 522	36	1 349	78	212	12
Total	264 111	2 770	36 894	1 463	77 720	2 807

Cuadro 1.4: Distribución del universo de estudiantes de quinto grado de secundaria

Departamento	Estatál				No estatal	
	Urbano		Rural			
	Estudiantes	IE	Estudiantes	IE	Estudiantes	IE
Amazonas	2 829	81	622	41	291	12
Ancash	6 909	174	2 130	93	1 645	59
Apurímac	4 163	77	348	19	216	10
Arequipa	11 321	151	635	20	5 281	183
Ayacucho	4 263	91	473	28	625	29
Cajamarca	7 778	151	3 998	244	1 423	62
Callao	6 394	41	0	0	2 685	133
Cusco	10 081	120	918	39	2 065	78
Huancavelica	3 446	95	1 308	66	130	10
Huánuco	5 157	75	770	33	656	24
Ica	7 166	74	264	16	1 199	48
Junín	8 492	92	1 760	48	2 469	102
La Libertad	10 812	162	1 146	91	2 936	118
Lambayeque	9 444	77	965	43	2 441	72
Lima	64 931	584	1 067	43	31 492	1 160
Loreto	5 205	51	560	35	1 011	22
Madre de Dios	673	9	27	3	31	2
Moquegua	1 691	30	245	9	204	8
Pasco	2 518	46	252	18	227	12
Piura	13 523	189	2 884	152	2 754	111
Puno	9 043	94	1 795	75	1 734	70
San Martín	5 578	88	505	23	283	12
Tacna	2 132	43	149	11	897	26
Tumbes	2 120	25	98	8	233	11
Ucayali	3 969	30	932	64	155	9
Total	209 638	2 650	23 851	1 222	63 083	2 383

5. Muestra efectiva

5.1. CUADROS CON LA DISTRIBUCIÓN DE LA MUESTRA

Cuadro 1.5: Distribución de la muestra efectiva de estudiantes de segundo grado de primaria

Departamento	Estatál												No estatal		
	Urbano						Rural								
	Unidocentes/ Multigrado			Polidocentes completos			Unidocentes/ Multigrado			Polidocentes completos					
	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE
	Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral	
Amazonas	40	40	2	90	88	3	62	63	10	23	21	1	37	37	2
Ancash	15	18	2	138	132	5	70	68	8	53	50	2	84	84	3
Apurímac	34	34	2	104	103	4	70	72	7	52	52	2	44	45	2
Arequipa	15	15	1	202	197	7	16	16	2	29	29	1	163	159	7
Ayacucho	47	46	2	119	116	4	61	58	7	15	11	1	50	50	2
Cajamarca	51	54	2	116	114	4	219	213	22	82	81	3	44	44	2
Callao				263	264	9							106	107	5
Cusco	40	39	2	142	137	5	173	173	12	59	59	2	80	79	3
Huancavelica	6	7	1	85	86	3	132	130	10	30	30	1	47	46	2
Huánuco	33	32	2	90	87	3	167	166	13	54	53	2	30	30	1
Ica	44	47	2	189	188	7	27	27	2	28	28	1	69	69	3
Junín	52	51	2	146	146	5	89	89	8	60	58	2	77	76	3
La Libertad	30	30	2	196	191	7	127	126	9	29	27	1	138	138	6
Lambayeque	223	224	11	2 615	2 534	93	464	462	47	435	427	17	426	426	22
Lima	62	55	3	835	836	30	17	16	2	27	29	1	991	992	46
Loreto	53	54	2	139	132	5	97	101	11	30	29	1	60	60	2
Madre de Dios	9	9	1	176	174	6	44	39	5	30	29	2	32	32	2
Moquegua	8	8	2	175	180	7	8	8	2	29	30	1	43	43	2
Pasco	21	22	2	113	105	4	75	76	8	17	17	1	36	36	2
Piura	45	43	2	216	211	8	98	99	10	45	44	2	121	122	5
Puno	29	29	2	148	146	5	65	65	8	42	43	2	45	44	2
San Martín	50	50	2	141	142	5	88	86	7	25	25	1	51	51	2
Tacna	28	25	1	230	213	8	5	5	1	42	36	2	58	58	2
Tumbes	9	8	2	219	213	8	8	8	1	20	19	1	53	50	2
Ucayali	48	45	2	144	145	5	69	69	7	29	28	1	57	58	2
Total	992	985	54	7 031	6 880	250	2 251	2 235	219	1 285	1 255	51	2 942	2 936	132

Cuadro 1.6: Distribución de la muestra efectiva de estudiantes de sexto grado de primaria

Departamento	Estatal												No estatal			
	Urbano						Rural									
	Unidocentes/ Multigrado			Polidocentes completos			Unidocentes/ Multigrado			Polidocentes completos						
	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE	
	Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		Lógico Matemática	Comunicación Integral		
Amazonas	41	40	2	88	83	3	50	50	10	22	22	1	33	32	2	
Ancash	13	13	2	140	139	5	67	68	8	46	45	2	72	72	3	
Apurímac	41	42	2	103	102	4	53	52	7	50	50	2	36	34	2	
Arequipa	11	11	1	201	202	7	19	19	2	29	28	1	154	156	7	
Ayacucho	43	42	2	108	108	4	44	44	5	22	21	1	44	44	2	
Cajamarca	41	41	2	118	120	4	161	159	20	53	53	3	38	38	2	
Callao	0	0	0	249	256	9	0	0	0	0	0	0	100	104	5	
Cusco	41	41	2	146	145	5	135	134	12	54	55	2	64	62	3	
Huancavelica	17	17	1	89	90	3	102	102	10	30	30	1	39	39	2	
Huánuco	33	33	2	90	89	3	70	70	10	60	60	2	30	29	1	
Ica	29	40	2	199	201	7	20	20	2	30	26	1	72	71	3	
Junín	59	59	2	148	148	5	52	50	8	55	55	2	70	71	3	
La Libertad	22	22	2	200	201	7	77	77	9	30	29	1	139	139	6	
Lambayeque	188	188	11	2 632	2 650	92	339	338	47	379	373	17	378	384	21	
Lima	61	59	3	861	870	30	15	15	2	30	30	1	904	924	46	
Loreto	48	48	2	142	144	5	67	65	11	30	29	1	60	60	2	
Madre de Dios	6	6	1	175	178	6	34	34	5	27	27	2	16	16	2	
Moquegua	8	8	2	185	182	7	10	10	2	30	30	1	40	38	2	
Pasco	25	24	2	113	116	4	65	65	8	15	15	1	49	48	2	
Piura	49	50	2	234	233	8	59	59	10	40	41	2	112	114	5	
Puno	20	20	2	147	146	5	76	78	8	55	56	2	39	39	2	
San Martín	24	25	2	145	146	5	36	36	7	27	27	1	39	39	2	
Tacna	28	29	1	213	204	8	5	5	1	41	37	2	55	58	2	
Tumbes	21	19	2	208	209	8	5	5	1	20	20	1	44	47	2	
Ucayali	34	33	2	147	147	5	38	38	7	27	28	1	59	58	2	
Total	903	910	54	7 081	7 109	249	1 599	1 593	212	1 202	1 187	51	2 686	2 716	131	

Cuadro 1.7: Distribución de la muestra efectiva de estudiantes de tercer grado de secundaria

Departamento	Estatal						No estatal		
	Urbano			Rural					
	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE
	Matemática	Comunicación		Matemática	Comunicación		Matemática	Comunicación	
Amazonas	368	371	14	53	52	4	76	75	4
Ancash	329	345	13	171	176	8	127	126	5
Apurímac	519	517	18	35	35	2	38	39	2
Arequipa	396	387	14	58	54	2	209	210	8
Ayacucho	443	437	16	32	32	2	78	80	3
Cajamarca	322	325	11	299	301	15	72	71	4
Callao	395	375	15	0	0	0	142	142	7
Cusco	466	456	16	80	76	3	78	75	3
Huancavelica	355	352	13	149	155	7	31	29	2
Huánuco	408	395	15	74	78	3	52	52	2
Ica	486	487	17	42	45	2	67	67	3
Junín	402	395	14	159	158	6	91	89	4
La Libertad	422	429	15	53	53	5	129	125	5
Lambayeque	422	417	16	102	104	4	98	95	4
Lima	928	922	33	76	76	4	1 080	1 040	48
Loreto	421	438	17	36	39	2	86	89	3
Madre de Dios	241	234	9	14	15	2	44	46	2
Moquegua	324	330	16	81	79	4	51	52	3
Pasco	459	442	17	35	33	3	48	45	2
Piura	409	397	14	202	204	10	167	163	6
Puno	436	436	15	147	154	6	85	87	3
San Martín	491	483	17	79	80	3	25	25	2
Tacna	385	353	15	31	28	2	128	118	5
Tumbes	479	480	18	26	25	2	77	71	3
Ucayali	433	430	16	81	82	4	49	52	2
Total	10 739	10 633	394	2 115	2 134	105	3 128	3 063	135

Cuadro 1.8: Distribución de la muestra efectiva de estudiantes de quinto grado de secundaria

Departamento	Estatal						No estatal		
	Urbano			Rural					
	Estudiantes		IE	Estudiantes		IE	Estudiantes		IE
	Matemática	Comunicación		Matemática	Comunicación		Matemática	Comunicación	
Amazonas	350	346	14	40	36	2	59	62	4
Ancash	312	336	13	162	161	8	114	121	5
Apurímac	439	461	17	22	23	2	33	34	2
Arequipa	382	389	14	59	59	2	213	208	8
Ayacucho	401	405	16	18	19	2	71	72	3
Cajamarca	291	297	11	258	260	15	70	69	4
Callao	375	387	14	0	0	0	126	129	7
Cusco	434	445	17	65	67	3	70	69	3
Huancavelica	326	311	13	89	88	7	30	27	2
Huánuco	388	375	15	55	53	3	35	36	2
Ica	447	468	17	39	37	2	68	67	3
Junín	393	395	14	148	146	6	86	89	4
La Libertad	393	401	15	44	51	5	110	112	5
Lambayeque	432	464	16	100	101	4	95	104	4
Lima	867	907	33	60	60	4	980	1 014	47
Loreto	404	433	17	37	34	2	79	76	3
Madre de Dios	195	200	9	9	8	2	16	19	2
Moquegua	333	332	16	67	65	4	48	49	3
Pasco	455	454	17	36	36	3	46	46	2
Piura	382	385	14	168	167	10	142	142	6
Puno	417	429	15	121	118	6	76	77	3
San Martín	441	451	17	73	76	3	53	52	2
Tacna	322	284	15	16	15	2	136	125	5
Tumbes	425	446	18	19	20	2	44	40	3
Ucayali	425	418	16	63	61	4	20	19	1
Total	10 029	10 219	393	1 768	1 761	103	2 820	2 858	133

5.2. NIVELES DE INFERENCIA Y REPRESENTATIVIDAD DE LA MUESTRA

La estratificación señalada en los cuadros anteriores se realizó con el fin de obtener una muestra representativa a escala nacional, mediante un diseño muestral de gran dispersión, lo que disminuye la varianza de las estimaciones a escala nacional. Sin embargo, el diseño de la muestra permite que los resultados sean también representativos para algunos estratos o uniones de estos. Así, podemos afirmar que la muestra es representativa en los siguientes niveles:

- Nacional
- Estatal / No estatal (según tipo de gestión)

Representativos solo del sector estatal:

- Urbano / Rural (según área de residencia)
- Polidocente completo, Unidocente / Polidocente multigrado (para las escuelas de primaria)

En cada uno de estos niveles de inferencia o dominios, el diseño muestral permite obtener estimaciones con una precisión del 5% a un nivel de confianza del 95%.

A nivel departamental, la muestra es referencial, es decir, se pueden comparar los resultados entre departamentos, pero estos no constituyen un nivel de inferencia por sí mismos.

5.3. EL CASO PARTICULAR DEL DEPARTAMENTO DE LAMBAYEQUE

Para el departamento de Lambayeque, en el diseño muestral para los grados de primaria, se realizó un sobremuestreo con el propósito de tener precisión en las estimaciones a nivel departamental y alguna precisión en los estimadores para los grandes estratos: Urbano/Rural, Estatal/No estatal; por lo que, a comparación de otros departamentos, la muestra de Lambayeque está conformada por aproximadamente 200 IE adicionales.

6. Estimaciones, márgenes de error y ponderaciones

A continuación, nos referiremos a algunos aspectos teóricos referidos al tipo de muestreo utilizado en la EN 2004 y cómo considerarlos para el cálculo de los errores muestrales de las estimaciones. Finalmente, presentaremos la forma como se han calculado las ponderaciones aplicadas a los estudiantes evaluados.

6.1. EFECTO DEL DISEÑO

En muestras de gran tamaño se asume normalidad en la distribución de las estimaciones de la población muestreada. Debido a que el diseño muestral involucra un muestreo en dos fases —en la primera fase se seleccionan IE y dentro de estas se seleccionan estudiantes—, se debe considerar el efecto del diseño (Kish 1965) en el cálculo del error estándar de las estimaciones.

El efecto del diseño, se define como la razón entre la varianza del estimador obtenida por el diseño muestral utilizado y la varianza del estimador correspondiente a un muestreo aleatorio simple. En un muestreo de dos fases, depende del tamaño del conglomerado

(β) y del coeficiente de correlación intraclase (ρ), como se aprecia en la siguiente ecuación:

$$\text{Efecto del diseño} = 1 + (\beta - 1) \rho$$

El coeficiente de correlación intraclase (ρ) proporciona una medida del grado de homogeneidad dentro de los conglomerados.

El tamaño de la muestra efectiva para un diseño muestral en dos fases es igual a:

$$n_{\text{efectivo}} = n^* \times \text{efecto del diseño}$$

donde n^* es el tamaño de muestra bajo un muestreo aleatorio simple

Por ejemplo, si asumiendo un muestreo aleatorio simple, la máxima dispersión en los resultados ($\rho = 0.5$) y un nivel de confianza del 95% ($Z=1.96$); se debe tener una muestra de 384 casos para obtener un margen de error máximo de ± 5 puntos porcentuales. Si se utiliza un muestreo bietápico, el tamaño de muestra que asegure los mismos resultados de confiabilidad y margen de error debe ser de $384 \times (\text{efecto de diseño})$. Para calcular el tamaño de muestra necesario para obtener estimaciones confiables con un margen de error pequeño —presentadas en el cuadro precedente—, se asumieron diferentes valores de β y de ρ para cada estrato.

En el siguiente cuadro se resume el tamaño de muestra mínimo para lograr los niveles de precisión y confianza deseados, junto con la muestra efectiva:

Cuadro 1.9: Tamaño de muestra mínimo y muestra efectiva de estudiantes para los diversos grados evaluados

Estrato		Tamaño de muestra mínimo	Muestra efectiva ⁸⁹			
			Segundo grado de primaria	Sexto grado de primaria	Tercer grado de secundaria	Quinto grado de secundaria
Nacional		5 000	14 291	13 471	15 830	14 617
Gestión	Estatal	3 000	11 355	10 785	12 767	11 797
	No estatal	2 500	2 936	2 686	3 063	2 820
Estatales						
Área	Urbano	3 300	7 865	7 984	10 633	10 029
	Rural	2 680	3 490	2 780	2 115	1 761
Característica	Polidocente completo	3 300	8 135	8 283		
	Polidocente Multigrado/Unidocente	2 085	3 220	2 502		

6.2. FÓRMULA PARA ESTIMAR EL ERROR ESTÁNDAR

Los valores que se presentan en este reporte se refieren a estimaciones puntuales,⁸⁹ calculadas a partir de las observaciones de la muestra. Debido a que se trata de observaciones realizadas en una muestra, estos valores corresponden a una muestra entre las

89. Por ejemplo, cuando se afirma que el 12,1% de los estudiantes de sexto de primaria alcanza el nivel suficiente en la prueba de Comprensión de textos escritos, estamos realizando una estimación puntual.

muchas que pudieron ser obtenidas a partir del mismo diseño muestral. Las estimaciones son pues variables aleatorias y por lo tanto están sujetas a una distribución probabilística. El resultado puntual no tiene valor en sí mismo, si no se sabe qué tan cerca se encuentra del verdadero valor poblacional. La única manera que se tiene de conocer cuál es la diferencia entre la estimación y el verdadero valor se da en términos probabilísticos. Para esto se debe calcular cuál es la variabilidad de las estimaciones, mediante el error estándar de las mismas.⁹⁰

En el diseño utilizado en la EN 2004, se deben tomar en cuenta para el cálculo de la varianza muestral dos aspectos: el uso de la estratificación y el uso de los conglomerados. Se debe calcular la varianza muestral dentro de cada estrato —tomando en cuenta los conglomerados— y luego calcular la varianza total como una suma ponderada de las varianzas de los estratos. A continuación se detallan las fórmulas para el cálculo de estas varianzas. Para esto se usarán las siguientes notaciones:

X_{ijh}	Puntajes del estudiante i de la IE j del estrato h
$\overline{X}_{jh}$	Media de los puntajes de la IE j del estrato h
$\overline{X}_h$	Media de las medias de los puntajes de las IE del estrato h
N	Número de IE en el universo
N_h	Número de IE en el universo en el estrato h
n_h	Número de IE en la muestra en el estrato h
M	Número de estudiantes en el universo
$\overline{M}_h$	Promedio de estudiantes por IE en el universo en el estrato h
$\overline{m}_h$	Promedio de estudiantes por IE en la muestra en el estrato h
L	Estratos

Varianza muestral dentro de cada estrato:

$$S_h^2 = (1 - f_{1h}) \frac{S_{Eh}^2}{n_h \overline{M}_h} + (1 - f_{2h}) \frac{S_{Dh}^2}{n_h \overline{m}_h}$$

donde

- S_{Eh}^2 es la varianza *entre* los conglomerados (IE) y S_{Dh}^2 es la varianza *dentro* de cada conglomerado
- $f_{1h} = \frac{n_h}{N_h}$ y $f_{2h} = \frac{\overline{m}_h}{\overline{M}_h}$ son las fracciones de muestreo de IE y de estudiantes respectivamente,

90. Por ejemplo, para la prueba de Comprensión de Textos escritos en sexto de primaria, el error estándar se ha estimado en 0,55. Esto nos permite construir intervalos de confianza, afirmando que hay un 95% de probabilidades de que el valor del parámetro en la población se encuentre entre 11,1% y 13,2%.

Varianza entre los conglomerados

$$S_{Eh}^2 = \bar{m}_h \frac{\sum_j^{n_h} (\bar{X}_{jh} - \bar{X}_h)^2}{n_h - 1}$$

Varianza dentro de cada conglomerado

$$S_{Dh}^2 = \frac{\sum_j^{n_h} \sum_i^{\bar{m}_h} (X_{ijh} - \bar{X}_{jh})^2}{n_h (\bar{m}_h - 1)}$$

Varianza total

La varianza total se calcula como la suma ponderada de las varianzas de cada estrato. El error estándar es la raíz cuadrada de la varianza total, es decir, $EE = \sqrt{S^2}$

$$S^2 = \sum_j^L W_h^2 S_h^2$$

donde

- $W_h = \frac{N_h \bar{M}_h}{M}$ son las ponderaciones de los estratos.

Si bien esta es la fórmula empleada para el cálculo del error estándar en un diseño muestral complejo —en los cuales se ha utilizado un muestreo por conglomerados y estratificación—, es de muy difícil aplicación.

6.3. MÉTODO DE REMUESTREO PARA ESTIMAR EL ERROR ESTÁNDAR

Para el cálculo de los errores estándar de este reporte se ha utilizado el paquete estadístico WesVar 4.2. Este programa informático calcula los errores mediante el uso de estimadores de remuestreo, como el denominado Jackknife N (JKn).

Este método se basa en información tomada de la muestra original (que actúa como el universo), a partir de la cual se generan varias muestras (llamadas réplicas) sobre las que se calcula el error estándar. El método JKn es utilizado cuando se trata de un muestreo estratificado y dos o más conglomerados son seleccionados en cada estrato. La fórmula que subyace a este algoritmo es la siguiente:

$$v(\hat{\theta}) = \sum_{g=1}^G (\hat{\theta}_g - \hat{\theta})^2$$

donde

- $\hat{\theta}$ es el estimador en la muestra original.
- $\hat{\theta}_g$ es el estimador en la muestra correspondiente a la g-ésima réplica
- G es el número de réplicas

El objetivo de esta técnica, como se puede apreciar, es estimar un valor en cada réplica, y luego calcular la varianza de todas las estimaciones, con lo que se estaría estimando de manera empírica —y no teórica, como en el caso de la estimación mediante la fórmula proporcionada en el punto anterior— la varianza del estimador.

6.4. FACTORES DE EXPANSIÓN Y PESOS

Con el fin de producir estimaciones válidas de la población total es necesario expandir las respuestas de cada uno de los individuos que conforman un estrato, de manera que la composición de la muestra por estrato sea semejante a la del universo bajo estudio. Para esto se calculan los factores de expansión. El factor de expansión es el recíproco de la probabilidad final de cada centro educativo y estudiante seleccionado.

En general, el factor de expansión de una muestra puede interpretarse como la cantidad de personas en la población que representa una persona en la muestra. Bajo el contexto de la EN 2004 y según el diseño muestral utilizado (bietápico, con probabilidad de selección de las UPM proporcionales al tamaño y estratificado) se debe aplicar un factor de expansión denominado *peso* a cada estudiante de la muestra, que depende del número de estudiantes que tiene la IE a la cual pertenece el alumno y el número de IE que tiene el estrato.

El peso en realidad *corrige* la distribución de la muestra en cada uno de los estratos para que concuerde con la distribución del universo. Así, se tiene que el peso de un estudiante, que pertenece al centro educativo i del estrato j está dado por:

$$\text{Factor de expansión} = \frac{\text{cantidad de alumnos que debía evaluarse en la IE}_i}{\text{cantidad de alumnos que se evaluó en la IE}_i} \times \frac{\text{población del estrato } j}{\Sigma \text{ alumnos que debían evaluarse en la IE de la muestra efectiva}}$$

Con el fin de que la muestra ponderada sume el total de casos de la muestra efectiva, se trabaja con el siguiente peso:

$$\text{Peso} = \text{Factor de expansión} \times \frac{\text{número de alumnos del estrato } j \text{ en la muestra efectiva}}{\text{cantidad de alumnos en el universo}}$$

De esta manera se garantiza que la muestra esté distribuida de manera proporcional, según los estratos que conforman la población objetivo.

Anexo 2: Diseño de cuadernillos de las pruebas aplicadas en la EN 2004

1. Aspectos generales

Uno de los objetivos de la EN 2004 era utilizar las pruebas de rendimiento para presentar el porcentaje de estudiantes en cada uno de los niveles de desempeño. Para esto se efectuó un muestreo representativo de los contenidos y competencias curriculares factibles de ser evaluados mediante una prueba estandarizada de lápiz y papel. Así se determinó que el número de ítemes con el que se proponía evaluar cada una de las áreas era bastante elevado; se necesitarían más de 100 preguntas para cada área, número que difícilmente podría ser resuelto por un estudiante en una sesión de una hora de duración.

Esta situación generó la necesidad de distribuir las preguntas en varios cuadernillos de manera que los estudiantes se enfrentaran a un número razonable de preguntas que sirvan para medir su desempeño de manera adecuada. Al respecto, González (2002) señala que

a fin de lograr evaluar amplios conocimientos de una disciplina en poco tiempo, frecuentemente se diseñan pruebas en las cuales se administran pocas preguntas a cada estudiante. Este tipo de diseño se llama diseño de pruebas rotadas o de matrices incompletas. Como resultado, existen varias formas de una misma prueba, cada forma tiene preguntas únicas y preguntas comunes a las otras formas, pero en conjunto se administra un número amplio de preguntas para cubrir una disciplina. (González: 10)

Esta estrategia es seguida en evaluaciones internacionales como TIMMS y PISA, en las que se aplica un conjunto de cuadernillos que son similares en lo que miden —pues deben basarse en la misma tabla de especificaciones— a fin de garantizar la equiparación de las puntuaciones y adecuada estimación de los parámetros de los ítems (Kolen y Brennan 2004).

Al efectuarse el análisis psicométrico con modelos de análisis Rasch, la medida de habilidad y dificultad de los ítemes se construye con una escala indeterminada, centrada para propósitos de identificación en la media aritmética de la dificultad de los ítemes o de la habilidad de las personas (Hambleton, Swaminathan y Rogers 1991).⁹¹ Es decir, dos cuadernillos aplicados a dos submuestras distintas de personas no necesariamente reportarán los parámetros de los ítemes o habilidades de las personas en la misma escala. Para ello, es necesario utilizar diversos diseños de escalamiento, a fin de poner todo en una métrica común (Navas 1996).⁹²

91. El evaluador es quien elige si centra las medidas en la dificultad de los ítemes o en la habilidad de las personas. En la EN 2004 se ha centrado la media en la habilidad de las personas.

92. Este proceso es análogo a la conversión de centímetros a pulgadas, o de grados centígrados a Fahrenheit, dependiendo del método de transformación utilizado.

Uno de los posibles diseños para colocar todo en una misma métrica es el llamado de ítemes comunes (Kolen y Brennan, 2004; Yu y Osborn 2005). Este método implica que las pruebas a equipararse comparten un conjunto de ítemes. Las respuestas a estos ítemes son utilizadas para realizar dicha equiparación, que en el caso de los modelos de Teoría de Respuesta al Ítem, generalmente consiste en poner los parámetros de los ítemes calculados en las diferentes pruebas en una métrica común. Es por esto que este procedimiento es muchas veces denominado *escalamiento* (Navas 1996).

Al haber un conjunto de ítemes en común, entre todos los cuadernillos se pueden poner fácilmente en una misma escala todos los parámetros de los ítems y estimaciones de la habilidad de las personas (Adams y Wu 2002). Para ello, se puede utilizar, entre otros procedimientos, el de calibración⁹³ concurrente: se analiza una sola matriz, programando al software para que lea los espacios en blanco como ítemes no aplicados (Kolen y Brennan 2004; Navas 1996). En estos casos se utilizan procesos de estimación de máxima verosimilitud conjunta (*joint maximun likelihood*) conocidos también como máxima verosimilitud incondicional (UCON), con matrices incompletas de datos como la que se muestra a continuación.

	Ítemes del Bloque 1					Ítemes del Bloque 2					Ítemes del Bloque 3					Ítemes del Bloque 4				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Alumnos que responden al cuadernillo 1	Patrón de respuestas individuales					Patrón de respuestas individuales														
Alumnos que responden al cuadernillo 2																				
Alumnos que responden al cuadernillo 3											Patrón de respuestas individuales					Patrón de respuestas individuales				
Alumnos que responden al cuadernillo 4																				

93. Calibrar: poner todos los parámetros de ítemes en una escala en común.

Si bien en el marco de la Teoría de Respuesta al Ítem existen otros métodos de calibración utilizando el diseño de ítemes comunes,⁹⁴ en algunas investigaciones (Cohen y Kim 1998; Hanson y Béguin 2002) se ha demostrado que el método de calibración conjunta produce mejores resultados frente a otros métodos de calibración de ítemes.

Wendler, Yu, Zeller y Carey (2004) señalan que los puntajes de los test pueden verse afectados por el orden en el cual aparecen los ítemes o bloques a los que se someten las personas examinadas. Esto se debería a dos factores principales que pueden afectar la estimación de los parámetros de los ítemes: la fatiga y la práctica en la resolución de la prueba.

A fin de controlar estos efectos, se utilizó la estrategia de bloques rotados. Es decir, los diferentes cuadernillos están formados por bloques que aparecen primero en una posición y luego en otra. Como se mostró en el ejemplo de la matriz incompleta, el cuadernillo 1 presenta en primer lugar al bloque 1 y luego al bloque 2; el cuadernillo 2 empieza por el bloque 2, que es seguido por el bloque 3. Este proceso continua hasta completar la espiral de rotaciones con el cuadernillo 4, que muestra en primer lugar al bloque 4 y en segundo lugar al bloque 1.⁹⁵ Así, cada cuadernillo comparte por lo menos un bloque con otros cuadernillos, apareciendo todos ellos conectados entre sí.

Las respuestas de todos los estudiantes son organizadas en una matriz incompleta de datos, y es reportado como parámetro de dificultad lo que podría considerarse un promedio de las dificultades al aparecer la pregunta en el primer y segundo bloque, lo que balancea los efectos de la fatiga y la práctica.

2. Descripción del procedimiento de armado de bloques y cuadernillos

En primer lugar, se eligió aquellas preguntas que mostraron un buen funcionamiento en la aplicación piloto. Para ello se consideró que estén espaciadas adecuadamente en sus estimaciones iniciales de dificultad y muestren un buen ajuste al modelo Rasch.

En segundo lugar, se dividió las preguntas en bloques, de manera que estos constituyan las partes para armar los cuadernillos. Como en la Evaluación Nacional 2004 se buscaba calcular el desempeño de los estudiantes, los bloques debían tener preguntas de varias capacidades. Este desempeño se calculó a partir de las diversas competencias adquiridas hasta el momento de la evaluación.

Respecto al número de preguntas por bloque, este varía dependiendo de la complejidad y extensión de preguntas que lo forman. Además se tomó en cuenta lo siguiente:

- El tiempo de duración de una sesión de aplicación, en cualquiera de las áreas, no debe exceder de una hora cronológica. Se estimó que el tiempo de duración para la solución de cada bloque debe ser de 25 a 30 minutos.
- Cumplir con el requerimiento señalado por Kolen y Brennan (2004) de tener por lo menos 20% de preguntas en común entre los cuadernillos para realizar la equiparación.

94. Por ejemplo, los métodos *mean/mean* y *mean/sigma* (Kolen y Brennan 2004).

95. Si bien en la matriz de datos parece que en el cuadernillo 4, primero se aplica el bloque 1 y luego el bloque 4, este ordenamiento solo es aplicable a la matriz de datos. La matriz se ordena según el número correlativo de los ítemes, no según el orden de aplicación.

Asimismo se tomaron ciertas consideraciones adicionales:

- Si bien lo recomendable es armar los cuadernillos empleando solo preguntas probadas en la aplicación piloto sin ningún tipo de modificaciones, se incluyeron, en una muy pequeña proporción preguntas nuevas o modificadas en los bloques, pues algunos ítemes no mostraron buen funcionamiento según los análisis efectuados en la aplicación piloto.
- Las preguntas modificadas o nuevas fueron distribuidas en los diversos bloques junto con preguntas ya probadas. No se armó ningún bloque solo con preguntas nuevas y/o modificadas.

3. Diseño de las pruebas de la EN 2004

3.1. PRUEBAS DEL ÁREA DE MATEMÁTICA

El diseño de las pruebas del área de Matemática aplicadas en la EN 2004 empleó dos bloques por cuadernillo con rotación en espiral. Se presenta en el siguiente cuadro un ejemplo del diseño de armado con ocho cuadernillos y ocho bloques.

Cuadro 2.1: Diseño de bloques rotados usado en las pruebas de rendimiento del área de Matemática

Cuadernillo	Parte 1	Parte 2
C01	Bloque 1	Bloque 2
C02	Bloque 2	Bloque 3
C03	Bloque 3	Bloque 4
C04	Bloque 4	Bloque 5
C05	Bloque 5	Bloque 6
C06	Bloque 6	Bloque 7
C07	Bloque 7	Bloque 8
C08	Bloque 8	Bloque 1

Como se puede apreciar en el cuadro anterior, la primera parte del cuadernillo 1 (C01) corresponde al bloque 1, y la segunda al bloque 2. En el cuadernillo 2 (C02) se rota el bloque 2, que aparece esta vez en la primera parte. En la segunda parte encontramos al bloque 3, que es rotado de posición en el siguiente cuadernillo. Este proceso continúa hasta que todos los bloques aparecen por lo menos una vez en la primera parte y otra en la segunda parte. Por lo tanto, se tendrá, como mínimo, tantos cuadernillos como bloques se haya construido.

3.2. PRUEBAS DE COMPRENSIÓN DE TEXTOS ESCRITOS

Estas pruebas fueron diseñadas considerando también la rotación de las posiciones de los bloques para aminorar el efecto del cansancio y la práctica. La particularidad de esta área es que los bloques no están constituidos por preguntas representativas de todos los contenidos, sino que cada bloque está constituido por un texto y sus respectivas preguntas. Además, los cuadernillos han sido contraídos en tres partes —evalúan la comprensión de tres textos— y se han rotado solo los textos que aparecen en la primera y la tercera posición.

Cuadro 2.2: Diseño de bloques rotados usado en las pruebas de rendimiento de Comprensión de textos escritos

Cuadernillo	Parte 1	Parte 2	Parte 3
C01	Texto 1	Texto 2	Texto 3
C02	Texto 3	Texto 4	Texto 5
C03	Texto 5	Texto 6	Texto 7
C04	Texto 7	Texto 8	Texto 9
C05	Texto 9	Texto 10	Texto 11
C06	Texto 11	Texto 12	Texto 13
C07	Texto 13	Texto 14	Texto 15
C08	Texto 15	Texto 16	Texto 1

Observamos que el cuadernillo 1 (C01) tiene como primera parte al texto 1, en la segunda aparece el texto 2 y en la tercera el texto 3. Así, el cuadernillo dos (C02) muestra en su primera parte al texto 3, en la segunda al texto 4 y en la tercera al texto 5. Este proceso continúa hasta que todos los textos de la primera y tercera parte han intercambiado por lo menos una vez sus posiciones.

Anexo 3: Proceso de corrección múltiple de los ítemes de producción de respuesta

1. Aspectos generales

Generalmente, cuando se realizan evaluaciones del rendimiento académico a gran escala, es común utilizar diferentes tipos de preguntas. Si bien en un inicio las evaluaciones nacionales utilizaron únicamente preguntas cerradas de opción múltiple, a partir de la EN 2001 se utilizaron también preguntas abiertas de producción de respuesta. En la actualidad, diversos sistemas de evaluación emplean este tipo de preguntas pues, como lo señalan Patz, Junker, Johnson y Mariano (2002), las respuestas abiertas son comúnmente utilizadas en las evaluaciones educativas a gran escala, pues permiten evaluar logros educativos más complejos.

Esta mejora en los instrumentos de evaluación trae consigo una complejidad adicional, y es que no pueden usarse sistemas automatizados para la corrección de las preguntas. Una prueba constituida solo por preguntas de opción múltiple puede ser calificada utilizando cualquier programa informático que permita comparar una plantilla de calificación y las respuestas de cada una de las personas evaluadas a cada una de las preguntas, a fin de asignarles el puntaje correspondiente. En cambio, la corrección de preguntas de producción de respuesta implica un procedimiento distinto. Generalmente, se trabaja con un conjunto de jueces⁹⁶ independientes que leen, evalúan y puntúan las respuestas dadas a las preguntas.

Al trabajarse con las valoraciones realizadas por los jueces, se encuentra siempre implicado algún grado de subjetividad, que se trata de reducir mediante la capacitación a los codificadores y la creación de manuales con criterios de calificación, que buscan homogenizar las evaluaciones realizadas por los jueces (Stemler 2004). Esto fue lo que se hizo para la calificación y transcripción de las respuestas de las pruebas aplicadas en la EN 2004. Los especialistas de la UMC se encargaron de efectuar el diseño de capacitación y de elaborar el respectivo manual de codificación de respuestas abiertas. Los codificadores se encontraban agrupados por área y grado, bajo la supervisión de un coordinador y los especialistas de la UMC. El proceso de corrección estaba organizado según los bloques que constituían las pruebas aplicadas en la EN 2004. Por ejemplo, los correctores eran capacitados en la corrección de todas las preguntas del bloque 03 y luego pasaban a corregir ese bloque en todos los cuadernillos que lo contenían. Posteriormente se recibía capacitación sobre las preguntas del bloque 04 y se pasaban a corregir los cuadernillos correspondientes. Se continuó así hasta completar la corrección de todos los bloques.

96. En algunos casos se refieren a las personas que realizan esta función, como calificadores, evaluadores, observadores, codificadores o correctores. Estos últimos términos fueron empleados en la EN 2004.

A pesar de lo importante y útil que puede ser la capacitación dada a los jueces, MacMillan (2000) advierte que por muy sistemático que haya sido el entrenamiento que se les ha dado, múltiples investigaciones demuestran que la variabilidad entre los correctores no puede ser eliminada del todo. Por ejemplo, Wolfe (2004) distingue tres tipos de efectos o sesgos que pueden presentar los jueces aún después de haber sido capacitados:

- Precisión / Imprecisión:

Implica qué tan bien ha sido entrenado o que tanta experiencia tiene un juez a fin de asignar calificaciones precisas. Es decir, se asume que hay un estándar, una calificación correcta, y se quiere ver qué tanto se aproxima un calificador a ese estándar.

La capacidad de asignar calificaciones precisas depende de múltiples factores, tanto personales (nivel educativo, entrenamiento, estilos de pensamiento, etc.) como contextuales (ausencia de distractores, interacciones sociales que se producen en el ambiente de calificación, etc.).

- Severidad / Leniencia

En estos casos puede haber calificadores que de forma consistente otorgan calificaciones más altas al compararse con el resto de calificadores (lenientes) o que otorgan menores calificaciones (severos).

- Centralidad / Extremismo

Este efecto implica que los calificadores suelen utilizar fundamentalmente las calificaciones intermedias (centralismo) o solo las altas y bajas (extremismo).

Dada la presencia de estos sesgos es muy importante evaluar la confiabilidad inter-evaluadores, que es definida por Stemler (2004) como el grado de acuerdo entre un conjunto particular de jueces, utilizando un instrumento de evaluación específico en un momento específico. Esta es una propiedad de la situación de evaluación, no del instrumento en sí mismo. En consecuencia deberá ser analizada cada vez que cambie la situación de evaluación.

Los análisis de la confiabilidad inter-evaluadores se han trabajado generalmente desde tres modelos teóricos: la Teoría Clásica de los Tests, la Teoría de la Generalizabilidad y el Modelo Rasch de Facetas Múltiples. Además, es importante considerar que la confiabilidad inter-evaluadores no es un concepto unitario pues existen diversas perspectivas respecto a su conceptualización y análisis. Así, Stemler (2004) propone una clasificación tripartita de los diversos tipos de estimaciones de la confiabilidad inter-evaluadores, cuya síntesis presentamos a continuación.

1. CONSENSO

Las estimaciones de consenso se basan en el supuesto de que dos o más observadores independientes deben llegar a un acuerdo exacto respecto de cómo aplicar los diversos niveles de una escala de calificación a un comportamiento observado.

Este tipo de índices es más útil cuando los datos son de naturaleza nominal, es decir, la escala de calificación representa diferencias cualitativas.

Los métodos más utilizados para este tipo de confiabilidad implican el cálculo del porcentaje de acuerdos, denominado en alguna ocasiones índice de acuerdos (Eскурra 1998), el coeficiente Kappa de Cohen y otros menos utilizados como la *J* de Jaccard, el índice *G* y el índice Delta propuesto por Martin y Femia (2004). Estos índices tienen la desventaja de tener que aplicarse a cada pregunta y a cada par de jueces.

Es importante señalar que existe una variante del coeficiente Kappa, conocida como Kappa múltiple, que se puede aplicar cuando hay más de dos evaluadores (Watkins 2002).

2. CONSISTENCIA

Las estimaciones de consistencia suponen que no es necesario que dos jueces muestren consenso en el uso de una escala de calificación, siempre y cuando las diferencias de consenso se apliquen de manera consistente. Es decir, puede ser que el juez A le asigne siempre o casi siempre la calificación 1 a cierto tipo de respuestas, mientras que el juez B les asigna siempre o casi siempre la calificación 3 a ese mismo tipo de respuestas. Por lo tanto, la diferencia de calificaciones entre ambos será predecible y podrá corregirse mediante el uso de constantes aditivas.

Este enfoque es utilizado cuando los datos son de naturaleza cuantitativa continua, pero también pueden aplicarse a variables ordinales, si se supone que representan un continuo a lo largo de una sola dimensión.

La ventaja de este enfoque es que si hay una consistencia en las calificaciones entre jueces, pueden aplicarse estrategias de corrección de las diferencias en severidad. Por ejemplo, si un juez A da una calificación un punto por encima de la de otro juez B y esto se repite de manera consistente, la correlación entre ambas calificaciones será bastante alta y se podrán equiparar las puntuaciones de ambos jueces restándoles un punto a todas las personas que fueron evaluadas por el juez A.

Los procedimientos empleados para las estimaciones de consistencia incluyen la correlación de Pearson (variables cuantitativas continuas) y de Spearman (variables ordinales). En los casos de tener varios jueces, se puede utilizar la *W* de Kendall (Cairns 2003; Legendre 2005).

3. MEDIDA

Las estimaciones de medida postulan que debe usarse toda la información disponible de los jueces cuando se trata de crear una calificación final para cada persona evaluada. Por ejemplo, se pueden controlar los efectos de severidad o lenidad al momento de asignar las calificaciones.

Se emplean cuando los diferentes niveles de la escala de calificación pretenden representar diferentes niveles de una variable latente unidimensional. También sirven cuando se tienen varios jueces y es imposible que todos los jueces califiquen todas las preguntas. Es decir, se tiene una matriz incompleta de datos, con conexiones por medio de personas comunes (una persona o conjunto de personas es calificada de forma independiente por diferentes jueces, pero no todas las personas son calificadas por todos los jueces).

El procedimiento más utilizado para este tipo de estimaciones es el análisis Rasch de Facetas Múltiples. En el caso de modelar las calificaciones de los jueces como una faceta, asume que la probabilidad de respuesta a una pregunta es una función de la habilidad

de quien responde a la pregunta, la dificultad de la pregunta y la severidad del corrector (Bond y Fox 2001). Otra alternativa es la propuesta por Patz, Junker, Johnson y Mariano (2002), conocida como Modelo Jerárquico de Evaluadores (HRM) que, dentro del marco de la Teoría de la Generalizabilidad, utiliza las distribuciones de las habilidades latentes propias de la Teoría de Respuesta al Ítem, en lugar de las distribuciones de puntajes verdaderos y los supuestos de distribución normal.

2. Diseño de análisis de la confiabilidad inter-evaluadores utilizado en la EN 2004

Solo los ítemes de producción de respuesta debían pasar por el proceso de corrección múltiple, pues solamente en ellos pueden intervenir diversos factores subjetivos del codificador. Como ya señalamos con anterioridad, la corrección de las pruebas aplicadas en la EN 2004 estaba organizada según los bloques de preguntas que las conformaban. Los codificadores eran capacitados en la corrección de las preguntas de un bloque y posteriormente pasaban a codificar las preguntas del mismo.

Se efectuó un diseño de análisis del proceso de corrección aplicando la corrección múltiple⁹⁷ en tres momentos distintos: al inicio, a la mitad y al final del mismo. El objetivo de estos tres momentos era poder retroalimentar sobre las discrepancias y deficiencias encontradas en las calificaciones y poder corregirlas durante el proceso de codificación, a fin de lograr cada vez una mayor homogeneidad y entendimiento en el uso de los criterios de calificación.

El primer momento se dio una vez finalizada la capacitación dada a los correctores, a fin de estimar los efectos del entrenamiento recibido. A partir de estos resultados se pudo reforzar algunos aspectos en aquellos correctores que presentaron algún tipo de sesgo en el uso de los criterios de calificación. El segundo momento ocurrió a la mitad del proceso de corrección, esperando que los efectos de la práctica mostraran un impacto positivo en el desempeño de los correctores. Finalmente, el tercer y último momento, se realizó cuando los correctores calificaban el último o penúltimo de los bloques. Tal y como esperábamos, la concordancia y precisión en el uso de los criterios de calificación, fue mejorando conforme avanzaba el proceso de corrección. Esta mejora se dio gracias a los efectos de la práctica y la oportuna retroinformación por parte de los especialistas de área a aquellos codificadores que no estaban empleando los criterios de calificación de forma precisa.

Para cada uno de los momentos de corrección múltiple se tomaron al azar 15 cuadernillos por cada área. Estos fueron corregidos de manera independiente por cada uno de los correctores, quienes pasaron sus respuestas a las respectivas fichas ópticas. Si bien los correctores podían consultar con sus coordinadores de grupo cada vez que tenían dudas respecto a la calificación de una pregunta, esta consulta no se efectuó para estos cuadernillos. Esta decisión se tomó porque se deseaba información sobre el uso de los criterios de corrección en su forma más pura posible y, si el evaluador le consultaba al coordinador, no se podría diferenciar durante los análisis aquellos efectos debidos a la decisión tomada por el corrector de aquellos debidos al criterio del coordinador. El proceso de

97. La corrección múltiple implica que una misma pregunta o conjunto de preguntas son evaluadas y puntuadas por cada uno de los calificadores de manera independiente.

lectura de fichas y depuración dio como resultado una matriz de datos, en la cual se tenía registrado, para cada uno de los cuadernillos, la codificación que le asignó cada corrector a cada una de las preguntas. Estas matrices fueron luego sometidas a un proceso de análisis de la confiabilidad inter-evaluadores.

3. Estrategias de análisis de la confiabilidad inter-evaluadores utilizado en la EN 2004

Las matrices de datos obtenidas en los tres momentos de corrección múltiple fueron sometidas a dos tipos de análisis: uno empleando los aportes de la Teoría Clásica de los Tests, que se centró en las estimaciones de consistencia, y otro basado en el modelo Rasch de facetas Múltiples, para realizar estimaciones de medida.

Los resultados de cada uno de los análisis se presentaron en un informe,⁹⁸ entregado a cada uno de los encargados de los grupos de codificadores. Este informe contenía una descripción del ajuste y consistencia global en el uso de las codificaciones. Luego se detallaban de manera precisa los índices de confiabilidad inter-evaluadores para cada uno de los ítemes, pudiendo así identificarse cuáles presentaban algún tipo de dificultad. Además se podía identificar claramente qué corrector estaba presentando alguna dificultad en el uso de las calificaciones y así retroinformarlo y hacerle un seguimiento más personalizado.

En cuanto a la codificación de las preguntas, por un lado, estas eran calificadas en diversas escalas, tanto de naturaleza dicotómica⁹⁹ como politómica¹⁰⁰, que tenían una naturaleza cuantitativa. Por otro lado, acompañaba a la calificación un segundo código, que tenía una naturaleza más cualitativa. Así un código «00» tenía semejanzas y diferencias con un código «01». Son semejantes en el sentido que ambos implican que la respuesta evaluada no recibe calificación alguna (cero puntos), pero se diferencian en la medida que «00» implica un tipo de error diferente al que implica el doble código «01». Este segundo código no ha influido en la construcción de la escala con el método Rasch. Sirvió para caracterizar el tipo de error, además de las estrategias de solución empleadas. Esta información luego fue usada en la elaboración de los Informes pedagógicos de resultados.

3.1. ANÁLISIS CLÁSICO Y DESCRIPTIVO

Estos análisis se efectuaron sobre la base de los dobles códigos, que tienen una naturaleza un tanto más cualitativa. Para ello se utilizó el coeficiente Kappa Múltiple (K), calculado para cada una de las preguntas mediante el programa informático MacKappa desarrollado por Watkins (2002). Este coeficiente nos permitió estudiar qué tanta concordancia existía entre los codificadores en cuanto al uso de los dobles códigos. La interpretación de la fuerza de la concordancia se efectuó según el siguiente cuadro:¹⁰¹

98. Se realizó un informe por cada área y grado evaluado.

99. Las preguntas recibían la calificación de cero (0) o uno (1).

100. Las escalas generalmente eran de tres puntos (calificación 0, 1 ó 2).

101. López de Ullibarri, I. y Pita, S. (1999).

Cuadro 3.1: Interpretación de la fuerza de la concordancia según el valor de K

Valor de K	Fuerza de la concordancia
< 0,20	Pobre
0,21 – 0,40	Débil
0,41 – 0,60	Moderada
0,61 – 0,80	Buena
0,81 – 1,00	Muy buena

Esta información además iba acompañada por un cuadro en el cual se muestra para cada uno de los cuadernillos calificados, la cantidad de codificadores que elegían cada uno de los códigos. Para los análisis descriptivos nos centramos en los cuadernillos que mostraban menos de un 70% de acuerdo en la calificación que debía recibir. Es decir, si un estudiante era calificado independientemente por 24 correctores, un acuerdo significativo implicaba que por lo menos 17 correctores coincidieran en el uso del doble código. Estas respuestas eran revisadas por los especialistas de cada área evaluada, los que intentaban determinar la causa de la baja concordancia y retroinformaban al grupo de codificadores para afinar los criterios.

A continuación se muestran algunos ejemplos de dichos cuadros con su respectiva interpretación:

Cuadro 3.2: Ejemplo 1, cantidad de correctores que emplea cada uno de los códigos

Estudiante	00	01	10	11
11725	1			24
14118	11	11	1	2

En este caso, vemos que se trata de 25 correctores. El estudiante 11725 recibió el código «11» por parte de 24 correctores, mientras que uno le dio la calificación «01». En cuanto el estudiante 14118, fue calificado con «00» por parte de 11 correctores, otros 11 le ponen «01», uno «10» y dos «11». Si bien se ve que hay muchas discrepancias, en este caso obedecen fundamentalmente al uso del segundo código.

Cuadro 3.3: Ejemplo 2, cantidad de correctores que emplea cada uno de los códigos

Estudiante	00	01	02	10	11	20	21
00201	1	12	10	13			
13905	1		31		1	3	

Podemos observar en este segundo ejemplo que ha habido dificultades para calificar las respuestas del estudiante 00201. Las discrepancias se dan no solo a nivel del primer dígito (puntuación) sino también a nivel del segundo código. En cuanto al estudiante 13905, la situación es mejor, pues 31 de los 36 codificadores coinciden en asignarle el código «10».

En general, como resultado de estos análisis se pudo observar que la concordancia entre los codificadores mejoró a medida que progresaba el proceso de calificación de las respuestas. Además, la mayoría de discrepancias obedecían al uso del segundo código. Esto no generó mayores problemas al momento de estimar la dificultad de la pregunta, pues este se basa solo en el primer dígito del doble código.¹⁰²

También se pudo apreciar gracias a este análisis que, en general, había mayor concordancia en el uso de los dobles códigos en el área de Matemática, frente al del área de Comunicación. Esto puede deberse a la naturaleza más precisa de las respuestas a las preguntas de Matemáticas, que se prestan a una menor subjetividad en su interpretación en relación a las de Comprensión de textos.

3.2. ANÁLISIS RASCH DE FACETAS MÚLTIPLES

Como este análisis de la consistencia inter-evaluadores efectúa una estimación de la medida del rasgo latente, solo se trabajó con el primer código, es decir, con el puntaje asignado por los codificadores a las respuestas evaluadas. Se utilizó el programa ConQuest¹⁰³ para el trabajo con este modelo, que busca estimar la probabilidad de una respuesta a una pregunta como una función de la dificultad de la pregunta, la habilidad de la persona que la responde y la severidad del juez que la califica. De esta manera se puede detectar la presencia de codificadores que se alejan significativamente del patrón típico de calificaciones, destacando por su severidad o lenidad.

La interpretación de los resultados se realizó mediante el índice de confiabilidad de la separación de codificadores, que puede tomar valores entre 0 y 1.¹⁰⁴ Cuanto más cercano a cero se encuentre el valor del índice, más homogéneas habrán sido las calificaciones de los correctores. Es decir, la interpretación de este índice sigue una lógica inversa en cuanto al significado común de los diversos índices de confiabilidad, pues mientras más alto es el valor, mayor separación existirá entre las calificaciones asignadas por los codificadores a las personas evaluadas, implicando ello menor confiabilidad. Estos resultados van acompañados con una prueba chi-cuadrado de igualdad de parámetros, que nos informa si las medidas estimadas para los codificadores presentan una diferencia mayor a aquella esperada por azar. Si los resultados de esta prueba estadística resultan significativos, por lo menos uno de los codificadores se diferencia del resto por ser más estricto o más leniente.

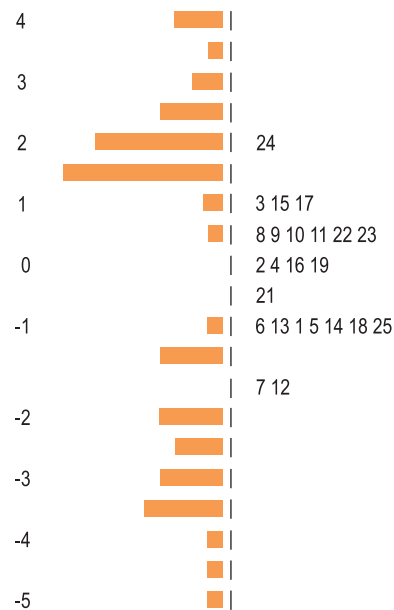
El diagnóstico de la severidad o lenidad se efectuó con ayuda de los mapas de codificadores. Como vemos en el ejemplo a continuación, en el extremo izquierdo aparece una escala numérica (va desde -5 a 4), que representa el grado de severidad de los correctores. Aparece también un gráfico de barras para simbolizar la distribución de la habilidad estimada para las personas evaluadas. Finalmente, cada uno de los codificadores se ubica en la escala de severidad según un número que sirve para identificarlo.

102. Es decir, la dificultad de las preguntas sólo toma en cuenta los aspectos cuantitativos de su calificación.

103. Wu, Adams y Wilson (1998).

104. Los resultados suelen presentarse con dos o tres decimales.

Gráfico 3.1: Ejemplo de un mapa de severidad / lenidad de los codificadores



En este mapa de codificadores, se puede observar que aquellos que muestran leniencia son los más alejados de la tendencia central que se ubican en el extremo inferior del mapa, es decir los calificadores 7 y 12. Mientras que los más severos han sido el 20 y el 24, que se encuentran en la parte superior del mapa.

Estos análisis se efectuaron tanto a nivel global (conjunto de preguntas) como para cada pregunta por separado. Además los resultados iban acompañados por una tabla que mostraba cuál fue la calificación que recibió cada uno de los estudiantes evaluados en una pregunta específica, por aquellos codificadores que se alejaban del patrón típico (moda) de calificaciones.

Cuadro 3.4: Calificaciones típicas (moda) asignadas a los diferentes estudiantes y calificaciones dadas por los codificadores más severos y más lenientes

Estudiante	7	12	Moda	20	24
05503	2	2	1	0	1
07318	2	2	1	0	1
16411	2	2	1	1	1
18101	1	2	0	0	0
25718	1	2	1	1	1
27514	2	2	2	1	1
29022	2	2	1	1	0
38413	2	1	2	1	1
40508	2	2	2	1	1
44011	2	1	1	0	1
46612	2	1	1	1	1
50130	2	1	2	1	1
54017	2	2	1	0	1
57205	0	0	0	1	0
61609	2	1	1	0	0

Se puede apreciar que los correctores 7 y 12 se han mostrado un poco más lenientes, mientras los correctores 20 y 24 han sido más severos que el resto de los correctores.

En términos generales, los resultados fueron bastante similares a los encontrados con las estimaciones de medida. Es decir, conforme fue progresando el proceso de codificación, se pudo observar una mejora en el uso de los criterios por parte de los calificadores. Esto fue gracias a que luego de cada corrección múltiple, los especialistas de área retroinformaban a los codificadores sobre su desempeño, lo que pretendía corregir cualquier imprecisión en el uso de los criterios de calificación.

Anexo 4: Establecimiento de puntos de corte

Introducción

El establecimiento de puntos de corte es el proceso por el cual se determinan cuáles son las tareas que los estudiantes enfrentados a una prueba deben realizar con éxito para evidenciar que han alcanzado un determinado criterio o nivel de dominio. Algunos métodos para establecer puntos de corte, llamados *ex ante*, permiten hacerlo previamente a la aplicación de los instrumentos a la población evaluada. Otros en cambio, los *ex post*, requieren que dicha aplicación ya se haya realizado pues emplean la data como elemento de análisis para el proceso.

En el caso de la EN 2004, se realizó una consulta *ex post* a jueces expertos en la materia (en pedagogía del área y currículo del grado). La estrategia metodológica asumida fue una adaptación de la desarrollada por CTB/MacGraw-Hill denominada Procedimiento Bookmark, la que también se utilizó para establecer los puntos de corte en la Evaluación Nacional 2001.

Establecer los puntos de corte de las pruebas ha permitido identificar el conjunto de tareas que debe contestar acertadamente un estudiante para poderlo clasificar en uno de los niveles de dominio definidos para el grado que cursa dicho estudiante. De esta manera, la población evaluada ha podido ser categorizada en función de los niveles de dominio definidos para la prueba y se ha obtenido el porcentaje de población que pertenece a cada uno de dichos niveles.

En el presente informe se detalla el método empleado para el establecimiento de los puntos de corte de la Evaluación Nacional 2004, así como su forma de aplicación.

1. La consulta de puntos de corte

1.1. FASES DEL TRABAJO

La consulta supuso las siguientes fases:

- a) Revisión bibliográfica, establecimiento de definiciones conceptuales y elaboración de materiales. En esta fase se establecieron los niveles de desempeño a utilizarse en cada grado. También se discutió el procedimiento estadístico a seguir y cada uno de los pasos que se deberían tener en cuenta durante la consulta
- b) Convocatoria a los expertos que cumplieran con los criterios de diversidad, experiencia en el nivel o grado y manejo del área consultada. Así se constituyeron grupos de trabajo para cada área consultada con la participación de docentes con experiencia en instituciones del nivel primario y secundario, públicas y privadas, de Lima y provincias. Estos tenían experiencia de aula en primaria y secundaria, o eran productores de materiales educativos para alguno de estos niveles. Asimismo,

mo, se congregó personal técnico de las instancias intermedias y central del MED con experiencia en el nivel primario y secundario.

1.2. METODOLOGÍA DEL PROCEDIMIENTO BOOKMARK

Como ya se señaló, dentro de la teoría y evaluación educacionales, esta estrategia metodológica es definida como una vía *ex post* para el establecimiento de puntos de corte, es decir, la determinación de los mismos es realizada después de que los estudiantes evaluados rinden la prueba. Esta característica es el supuesto central del procedimiento, ya que trabaja sobre la base del ordenamiento de los ítemes evaluados de acuerdo con el parámetro de dificultad calculado mediante el modelo Rasch.¹⁰⁵

Los jueces convocados analizan los ítemes, ordenados crecientemente por dificultad, y definen qué hace a cada uno más difícil que el anterior, de tal manera que se especifican los procesos y los contenidos involucrados en cada uno y se identifica la secuencia gradual de dificultad, construida a partir de los resultados de los estudiantes evaluados.

Una vez que los jueces han concluido con el análisis de los ítemes, se procede a hacer una descripción de los niveles de dominio evaluados en la prueba. Estos niveles de dominio agrupan una serie de tareas que tienen en común cierto grado de dificultad en función de los contenidos y de las capacidades que involucran. Cada juez, entonces, estima bloques o conjuntos de preguntas que corresponden a los niveles de dominio mencionados anteriormente. Si los niveles de dominio definidos para la prueba son n , cada juez llega a establecer « $n+1$ » segmentos de ítemes por medio de la ubicación de n marcas que distinguen los ítemes que corresponden a cada nivel.

En nuestro caso, se trabajó con tres niveles de dominio: suficiente, básico y previo. Entonces los jueces distinguieron, en primer lugar, a los ítemes que corresponden al nivel suficiente de los que están por encima del nivel suficiente (marca 1); en segundo lugar, se distinguieron los del nivel suficiente de los del nivel básico (marca 2); y, en tercer lugar, se distinguieron los del nivel básico de los del previo (marca 3). Este momento inicial de clasificación de los ítemes es denominado «primera ronda de clasificación». Cabe señalar que, en esta primera ronda, cada juez trabaja de manera individual, es decir, sin conocer dónde están colocando sus marcas los otros jueces.

Esta primera clasificación efectuada por los jueces es ingresada a la computadora y el resultado que se obtiene es el cuadro de presentación de resultados,¹⁰⁶ que expresa la coincidencia o no de opinión técnica por parte de los jueces en relación a los ítemes.¹⁰⁷ Estos son considerados dentro de un nivel determinado cuando el 100% de los jueces involucrados coincide en tal clasificación. En los otros casos, los ítemes son materia de debate entre los jueces. En dichos debates los jueces argumentan las razones por las que han clasificado al ítem en uno u otro nivel. El debate no busca forzar el acuerdo,

105. Los ítemes ordenados están contenidos en un recurso denominado cuadernillo Rasch.

106. Este condensa el veredicto de los jueces y muestra sobre qué ítemes no hay duda, puesto que los jueces concurren en clasificarlos dentro de un determinado nivel y, a la vez, especifica aquellos sobre los cuales hay necesidad de que cada quien argumente sobre su ubicación. Este proceso corresponde al momento previo de una segunda ronda de clasificación de dichos ítemes.

107. Cabe resaltar que el término «coincidencia o no de opinión técnica» es muy importante dentro de este procedimiento, ya que cada juez emite un veredicto sin que medie proceso alguno de búsqueda de consenso o de convencimiento entre jueces.

simplemente se intenta dar más elementos de juicio para que los jueces reflexionen. Después de la revisión de los ítemes sobre los cuales no hay coincidencia, los jueces vuelven a clasificarlos de manera individual y a registrar en el formato de clasificación el código respectivo en la columna de la segunda ronda.¹⁰⁸

El cuadro resultante de la segunda ronda de clasificación también es mostrada a los jueces, pero, a diferencia del primer cuadro, este contiene el porcentaje de alumnos que estaría ubicado dentro de cada bloque de preguntas, en otras palabras, que correspondería a cada nivel de dominio. Este resultado es denominado «impacto de la data». Los jueces, a partir de dicha distribución porcentual, reaccionan y comentan los factores que estarían interviniendo en la obtención de esa distribución.

Después del momento de intercambio de opiniones generado a partir del impacto de la data, los jueces vuelven a clasificar los ítemes, utilizando la tercera columna del formato de clasificación. Este momento es denominado «tercera ronda». En esta tercera clasificación, cada juez puede o no tomar en cuenta la distribución observada de estudiantes así como las razones dadas por los demás jueces para explicar la misma. Una vez clasificados los ítemes, los resultados son tabulados como las dos veces anteriores, pero aplicando un nivel menor de exigencia al margen de coincidencia: 80%; es decir, los ítemes son clasificados dentro de un determinado nivel de dominio si, por lo menos, el 80% de jueces coincide respecto a esta clasificación del ítem en particular. Los jueces evalúan el proceso desarrollado una vez finalizada la tercera ronda y recogidos los formatos de clasificación.

108. Los ítemes sobre los cuales hubo coincidencia de opinión en la primera ronda mantienen el código original, no cambian.

Anexo 5: Tablas de resultados generales

Segundo grado de primaria

COMUNICACIÓN INTEGRAL

Cuadro 5.1: Porcentaje de estudiantes según nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e. ¹⁰⁹
Suficiente	15,1	0,64
Básico	23,7	0,71
Previo	15,1	0,55
< Previo	46,1	1,27

Cuadro 5.2: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z ¹¹⁰	p ¹¹¹
	%	e.e.	%	e.e.		
Suficiente	15,1	0,80	15,2	0,84	-0,09	0,931
Básico	24,5	0,87	22,9	0,86	1,31	0,191
Previo	16,3	0,67	13,9	0,70	2,48	0,013
< Previo	44,2	1,42	48,0	1,51	-1,83	0,067

Cuadro 5.3: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	9,5	0,58	48,5	2,24	- 16,9	0,000
	21,9	0,78	34,7	1,51	-7,5	0,000
	15,9	0,62	10,2	0,80	5,6	0,000
	52.7	1.43	6.6	0.98	26.6	0.000

109. Es el error estándar de estimación para el porcentaje.

110. Se realizó una prueba Z para diferencia de proporciones, comparando ambos estratos en cada uno de los niveles de desempeño.

111. Es el índice p asociado a la distribución Z. Con un 95% de confianza, se dice que los valores inferiores a 0,05 resultan significativos; mientras que con un 99% de confianza, los valores inferiores a 0,01 son significativos.

Cuadro 5.4: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente	14,5	0,91	2,5	0,42	11,97	0,000
Básico	30,4	1,06	9,8	0,98	14,27	0,000
Previo	18,9	0,78	11,7	0,99	5,71	0,000
< Previo	36,2	1,85	76,0	1,94	-14,85	0,000

Cuadro 5.5: Porcentaje de estudiantes por tipo de institución educativa según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Polidocentes		Multigrado		Z	p
	%	e.e.	%	e.e.		
Suficiente	13,9	0,85	1,7	0,33	13,38	0,000
Básico	29,5	1,01	8,1	0,96	15,36	0,000
Previo	18,8	0,78	10,8	0,97	6,43	0,000
< Previo	37,7	1,82	79,4	1,75	-16,52	0,000

Cuadro 5.6: Porcentaje de estudiantes por departamento según nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	2,7	1,0
	Básico	17,8	4,0
	Previo	22,9	4,5
	Debajo de Previo	56,6	6,9
ÁNCASH	Suficiente	10,9	3,2
	Básico	21,9	5,0
	Previo	12,7	2,0
	Debajo de Previo	54,6	8,1
APURÍMAC	Suficiente	2,3	1,8
	Básico	6,8	4,3
	Previo	6,3	3,3
	Debajo de Previo	84,7	9,2
AREQUIPA	Suficiente	31,7	5,7
	Básico	34,8	4,0
	Previo	14,4	3,1
	Debajo de Previo	19,1	5,0
AYACUCHO	Suficiente	4,6	2,7
	Básico	13,6	5,8
	Previo	7,2	2,9
	Debajo de Previo	74,6	10,6
CAJAMARCA	Suficiente	5,1	1,3
	Básico	19,1	3,0
	Previo	18,2	2,9
	Debajo de Previo	57,6	5,8
CALLAO	Suficiente	26,9	5,5
	Básico	40,1	3,5
	Previo	15,6	2,6
	Debajo de Previo	17,4	5,4
CUSCO	Suficiente	3,3	1,8
	Básico	10,4	3,1
	Previo	10,1	2,9
	Debajo de Previo	76,2	6,5
HUANCAVELICA	Suficiente	2,8	1,7
	Básico	12,0	3,9
	Previo	17,3	4,0
	Debajo de Previo	67,9	6,9
HUÁNUCO	Suficiente	6,5	3,4
	Básico	12,6	3,7
	Previo	15,4	3,3
	Debajo de Previo	65,5	8,5
ICA	Suficiente	20,3	4,7
	Básico	33,1	3,7
	Previo	20,0	3,1
	Debajo de Previo	26,6	6,5
JUNÍN	Suficiente	12,1	3,3
	Básico	25,8	3,7
	Previo	19,6	3,6
	Debajo de Previo	42,6	7,0

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	20,5	3,6
	Básico	26,9	3,3
	Previo	16,1	2,6
	Debajo de Previo	36,5	6,8
LAMBAYEQUE	Suficiente	18,1	1,4
	Básico	28,5	1,0
	Previo	17,3	0,8
	Debajo de Previo	36,1	1,8
LIMA	Suficiente	31,5	1,9
	Básico	36,1	1,6
	Previo	15,4	0,9
	Debajo de Previo	17,0	2,2
LORETO	Suficiente	2,0	1,0
	Básico	6,9	2,3
	Previo	7,3	1,8
	Debajo de Previo	83,8	4,0
MADRE DE DIOS	Suficiente	10,7	3,2
	Básico	29,1	4,8
	Previo	20,7	2,2
	Debajo de Previo	39,5	7,3
MOQUEGUA	Suficiente	34,0	7,2
	Básico	38,0	3,9
	Previo	15,8	4,0
	Debajo de Previo	12,3	3,6
PASCO	Suficiente	7,7	3,8
	Básico	15,5	4,0
	Previo	20,2	3,8
	Debajo de Previo	56,7	8,2
PIURA	Suficiente	11,3	2,9
	Básico	23,0	3,4
	Previo	14,5	2,3
	Debajo de Previo	51,3	6,1
PUNO	Suficiente	6,2	2,7
	Básico	18,0	4,3
	Previo	20,2	3,0
	Debajo de Previo	55,6	7,2
SAN MARTÍN	Suficiente	7,6	3,9
	Básico	16,8	5,2
	Previo	13,1	3,4
	Debajo de Previo	62,5	9,7
TACNA	Suficiente	33,2	5,2
	Básico	44,7	3,6
	Previo	13,0	2,5
	Debajo de Previo	9,1	2,6
TUMBES	Suficiente	11,5	3,2
	Básico	32,4	5,1
	Previo	18,7	1,9
	Debajo de Previo	37,4	6,3
UCAYALI	Suficiente	2,7	1,3
	Básico	10,7	3,1
	Previo	17,2	3,3
	Debajo de Previo	69,3	6,4

Cuadro 5.7: Porcentaje de estudiantes dentro de cada nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	9,6	0,49
Básico	27,4	0,78
Previo	18,8	0,59
< Previo	44,2	1,22

Cuadro 5.8: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	10,9	0,66	8,3	0,60	2,91	0,004
	29,8	0,97	24,8	0,91	3,76	0,000
	18,3	0,72	19,4	0,75	-1,06	0,290
	40,9	1,37	47,5	1,47	-3,28	0,001

Cuadro 5.9: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente	6,1	0,43	30,4	1,97	-12,1	0,000
Básico	24,0	0,86	47,4	1,36	-14,5	0,000
Previo	19,5	0,66	15,0	1,12	3,5	0,001
< Previo	50,4	1,38	7,2	0,97	25,6	0,000

Cuadro 5.10: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	8,8	0,65	2,4	0,46	8,04	0,000
	32,8	1,15	11,5	1,12	13,27	0,000
	24,1	0,85	12,9	0,98	8,63	0,000
	34,3	1,70	73,3	1,97	-14,99	0,000

Cuadro 5.11: Porcentaje de estudiantes por tipo de institución educativa según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Polidocentes		Multigrado		Z	p
	%	e.e.	%	e.e.		
Suficiente	8,4	0,61	2,2	0,48	7,99	0,000
Básico	31,9	1,12	9,9	1,02	14,52	0,000
Previo	23,9	0,84	11,6	0,93	9,81	0,000
< Previo	35,9	1,7	76,4	1,78	-16,45	0,000

**Cuadro 5.12: Porcentaje de estudiantes por departamento
según nivel de desempeño**

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	9,4	4,2
	Básico	18,9	3,8
	Previo	17,0	3,5
	Debajo de Previo	54,7	7,5
ÁNCASH	Suficiente	6,4	2,2
	Básico	24,7	4,7
	Previo	18,5	3,5
	Debajo de Previo	50,4	8,4
APURÍMAC	Suficiente	3,2	2,5
	Básico	7,6	4,2
	Previo	8,7	2,9
	Debajo de Previo	80,4	9,0
AREQUIPA	Suficiente	18,8	4,6
	Básico	43,6	3,9
	Previo	19,2	3,1
	Debajo de Previo	18,4	4,6
AYACUCHO	Suficiente	3,5	2,5
	Básico	15,2	5,4
	Previo	13,1	3,1
	Debajo de Previo	68,2	9,3
CAJAMARCA	Suficiente	4,5	1,2
	Básico	20,7	3,6
	Previo	20,1	2,7
	Debajo de Previo	54,7	5,4
CALLAO	Suficiente	16,7	4,6
	Básico	41,8	4,2
	Previo	25,3	3,5
	Debajo de Previo	16,2	4,9
CUSCO	Suficiente	2,0	1,0
	Básico	12,7	3,0
	Previo	14,2	2,8
	Debajo de Previo	71,0	5,6
HUANCAVELICA	Suficiente	1,7	1,2
	Básico	15,9	4,3
	Previo	21,9	3,1
	Debajo de Previo	60,4	7,0
HUÁNUCO	Suficiente	4,6	1,8
	Básico	16,1	5,3
	Previo	13,1	3,6
	Debajo de Previo	66,3	8,3
ICA	Suficiente	11,4	3,2
	Básico	41,9	5,4
	Previo	18,9	2,6
	Debajo de Previo	27,7	6,8

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	11,0	2,3
	Básico	30,8	3,4
	Previo	22,5	2,6
	Debajo de Previo	35,7	6,2
LAMBAYEQUE	Suficiente	11,4	1,1
	Básico	33,5	1,3
	Previo	21,2	0,9
	Debajo de Previo	33,9	1,7
LIMA	Suficiente	19,1	1,5
	Básico	42,3	1,7
	Previo	21,6	1,4
	Debajo de Previo	17,1	1,7
LORETO	Suficiente	2,4	1,5
	Básico	7,6	2,2
	Previo	9,5	2,0
	Debajo de Previo	80,5	4,5
MADRE DE DIOS	Suficiente	5,8	2,0
	Básico	31,6	5,4
	Previo	22,5	1,9
	Debajo de Previo	40,1	6,6
MOQUEGUA	Suficiente	25,7	5,9
	Básico	45,4	4,1
	Previo	20,2	4,4
	Debajo de Previo	8,7	3,0
PASCO	Suficiente	3,4	1,3
	Básico	18,3	4,8
	Previo	24,1	3,6
	Debajo de Previo	54,3	6,8
PIURA	Suficiente	4,1	1,3
	Básico	24,5	3,7
	Previo	20,5	2,8
	Debajo de Previo	50,9	5,9
PUNO	Suficiente	3,9	1,6
	Básico	22,7	4,4
	Previo	21,5	3,9
	Debajo de Previo	51,9	8,4
SAN MARTÍN	Suficiente	4,9	2,7
	Básico	18,2	4,9
	Previo	14,8	4,0
	Debajo de Previo	62,2	9,7
TACNA	Suficiente	29,3	3,8
	Básico	43,2	4,7
	Previo	20,5	4,6
	Debajo de Previo	7,0	2,6
TUMBES	Suficiente	9,5	3,1
	Básico	33,2	5,8
	Previo	17,9	2,7
	Debajo de Previo	39,5	7,9
UCAYALI	Suficiente	1,4	0,8
	Básico	13,5	5,1
	Previo	14,1	3,6
	Debajo de Previo	71,0	8,2

Sexto grado de primaria

COMUNICACIÓN INTEGRAL

Cuadro 5.13: Porcentaje de estudiantes según nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	12,1	0,55
Básico	28,1	0,75
Previo	35,7	0,74
< Previo	24,1	1,05

Cuadro 5.14: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	11,0	0,65	13,3	0,76	-2,30	0,021
	28,1	0,93	28,2	0,94	-0,08	0,940
	36,2	0,92	35,1	1,00	0,81	0,418
	24,7	1,24	23,5	1,21	0,69	0,489

Cuadro 5.15: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	8,2	0,46	36,1	2,04	-13,3	0,000
	26,0	0,85	40,9	1,24	-9,9	0,000
	38,3	0,81	19,4	1,54	10,9	0,000
	27.5	1.20	3.5	0.60	17.9	0.000

Cuadro 5.16: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	11,0	0,64	2,4	0,42	11,23	0,000
	32,8	1,07	11,8	1,13	13,49	0,000
	40,2	0,91	34,4	1,60	3,15	0,002
	16,0	1,16	51,5	2,47	-13,01	0,000

Cuadro 5.17: Porcentaje de estudiantes por tipo de institución educativa según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Polidocentes		Multigrado		Z	p
	%	e.e.	%	e.e.		
Suficiente	10,5	0,60	1,7	0,38	12,39	0,000
Básico	31,6	1,03	10,5	1,04	14,42	0,000
Previo	39,8	0,90	34,3	1,70	2,86	0,004
< Previo	18,1	1,25	53,5	2,33	-13,39	0,000

Cuadro 5.18: Porcentaje de estudiantes por grupo de pobreza según nivel de desempeño

Niveles de desempeño	IE Estatales urbanas					
	Pobres		No pobres		Z	p
	%	e.e.	%	e.e.		
Suficiente	5,9	0,66	13,9	0,84	-7,49	0,000
Básico	25,6	1,80	36,9	1,13	-5,32	0,000
Previo	41,7	1,45	39,3	1,12	1,31	0,190
< Previo	26,8	2,17	10,0	0,85	7,21	0,000

Cuadro 5.19: Porcentaje de estudiantes por departamento según nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	3,7	1,5
	Básico	12,0	2,8
	Previo	48,8	5,1
	Debajo de Previo	35,5	6,7
ÁNCASH	Suficiente	11,7	2,4
	Básico	27,0	4,3
	Previo	38,1	4,0
	Debajo de Previo	23,2	6,4
APURÍMAC	Suficiente	1,2	0,7
	Básico	14,3	6,0
	Previo	30,9	5,9
	Debajo de Previo	53,7	10,9
AREQUIPA	Suficiente	20,0	3,5
	Básico	42,4	3,6
	Previo	30,3	3,3
	Debajo de Previo	7,3	2,1
AYACUCHO	Suficiente	5,3	1,8
	Básico	12,9	6,2
	Previo	39,6	3,1
	Debajo de Previo	42,2	8,2
CAJAMARCA	Suficiente	4,8	1,4
	Básico	17,6	3,9
	Previo	39,1	3,7
	Debajo de Previo	38,6	6,3
CALLAO	Suficiente	20,9	3,0
	Básico	36,1	2,1
	Previo	31,7	2,7
	Debajo de Previo	11,3	2,5
CUSCO	Suficiente	3,0	1,6
	Básico	15,3	3,5
	Previo	33,8	3,8
	Debajo de Previo	47,9	6,6
HUANCAVELICA	Suficiente	4,1	2,4
	Básico	15,4	6,3
	Previo	39,4	5,9
	Debajo de Previo	41,2	8,3
HUÁNUCO	Suficiente	6,8	2,9
	Básico	18,6	4,5
	Previo	36,0	4,6
	Debajo de Previo	38,6	9,5
ICA	Suficiente	14,2	3,7
	Básico	33,5	3,3
	Previo	42,3	4,2
	Debajo de Previo	10,0	2,7
JUNÍN	Suficiente	7,0	2,2
	Básico	31,3	5,4
	Previo	37,6	4,3
	Debajo de Previo	24,2	6,5

Cont. Cuadro 5.19

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	15,6	3,6
	Básico	27,0	2,2
	Previo	41,4	3,5
	Debajo de Previo	16,0	4,2
LAMBAYEQUE	Suficiente	13,2	0,9
	Básico	32,5	1,1
	Previo	38,2	1,2
	Debajo de Previo	16,1	1,1
LIMA	Suficiente	21,6	1,4
	Básico	40,5	1,5
	Previo	31,8	1,6
	Debajo de Previo	6,1	0,9
LORETO	Suficiente	2,0	0,8
	Básico	16,0	3,6
	Previo	33,0	3,8
	Debajo de Previo	49,0	7,3
MADRE DE DIOS	Suficiente	11,2	2,7
	Básico	33,6	4,0
	Previo	39,3	5,5
	Debajo de Previo	15,9	2,6
MOQUEGUA	Suficiente	19,1	5,0
	Básico	41,8	4,3
	Previo	32,0	4,5
	Debajo de Previo	7,2	2,2
PASCO	Suficiente	8,3	2,4
	Básico	25,6	4,6
	Previo	40,0	4,0
	Debajo de Previo	26,1	6,2
PIURA	Suficiente	8,9	2,0
	Básico	30,6	2,9
	Previo	36,9	3,7
	Debajo de Previo	23,6	5,7
PUNO	Suficiente	6,8	3,4
	Básico	16,3	4,0
	Previo	34,8	4,1
	Debajo de Previo	42,0	7,0
SAN MARTÍN	Suficiente	8,1	4,3
	Básico	18,9	5,6
	Previo	35,2	5,7
	Debajo de Previo	37,8	10,2
TACNA	Suficiente	23,9	3,8
	Básico	36,2	3,3
	Previo	33,6	4,1
	Debajo de Previo	6,3	2,0
TUMBES	Suficiente	5,0	1,9
	Básico	30,0	3,3
	Previo	45,3	2,0
	Debajo de Previo	19,7	3,1
UCAYALI	Suficiente	4,6	1,8
	Básico	15,7	2,6
	Previo	39,6	4,9
	Debajo de Previo	40,1	5,3

Cuadro 5.20: Porcentaje de estudiantes por nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	7,9	0,49
Básico	34,7	0,97
Previo	12,7	0,46
< Previo	44,8	1,28

Cuadro 5.21: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	9,0	0,68	6,8	0,55	2,52	0,012
	36,4	1,16	32,9	1,19	2,11	0,035
	12,7	0,57	12,6	0,65	0,12	0,908
	41,9	1,42	47,7	1,56	-2,75	0,006

Cuadro 5.22: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente	4,4	0,36	29,7	2,18	-11,5	0,000
Básico	31,9	1,10	51,7	1,47	-10,8	0,000
Previo	13,6	0,52	7,0	0,74	7,3	0,000
< Previo	50,4	1,42	11,6	1,27	20,3	0,000

Cuadro 5.23: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	5,8	0,51	1,3	0,31	7,54	0,000
	39,5	1,37	16,0	1,62	11,08	0,000
	15,0	0,65	10,6	0,86	4,08	0,000
	39.7	1.75	72.1	2.17	11.62	- 0.000

Cuadro 5.24: Porcentaje de estudiantes por tipo de institución educativa según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Polidocentes		Multigrado		Z	p
	%	e.e.	%	e.e.		
Suficiente	5,6	0,48	0,9	0,25	8,68	0,000
Básico	38,1	1,32	14,7	1,68	10,91	0,000
Previo	14,7	0,61	10,4	0,98	3,73	0,000
< Previo	41,6	1,69	74,0	2,25	-11,51	0,000

Cuadro 5.25: Porcentaje de estudiantes por grupo de pobreza según nivel de desempeño

Niveles de desempeño	IE Estatales urbanas					
	Pobres		No pobres		Z	p
	%	e.e.	%	e.e.		
Suficiente	2,3	0,40	7,8	0,67	-7,05	0,000
Básico	29,8	1,82	44,9	1,58	-6,27	0,000
Previo	14,0	1,07	15,5	0,79	-1,13	0,259
< Previo	53,9	2,29	31,8	1,90	7,43	0,000

Cuadro 5.26: Porcentaje de estudiantes por departamento según nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	1,1	0,8
	Básico	21,1	4,9
	Previo	16,6	3,4
	Debajo de Previo	61,2	6,0
ÁNCASH	Suficiente	6,9	2,3
	Básico	30,9	4,5
	Previo	13,6	2,9
	Debajo de Previo	48,7	6,7
APURÍMAC	Suficiente	0,7	0,6
	Básico	11,4	5,6
	Previo	7,3	1,6
	Debajo de Previo	80,6	6,2
AREQUIPA	Suficiente	15,0	3,7
	Básico	49,5	3,9
	Previo	12,3	3,0
	Debajo de Previo	23,2	3,7
AYACUCHO	Suficiente	2,6	1,4
	Básico	16,9	5,0
	Previo	9,7	2,3
	Debajo de Previo	70,8	7,8
CAJAMARCA	Suficiente	3,8	1,4
	Básico	32,8	5,3
	Previo	16,1	2,7
	Debajo de Previo	47,4	6,5
CALLAO	Suficiente	9,0	2,5
	Básico	55,0	3,3
	Previo	13,0	2,3
	Debajo de Previo	23,0	4,4
CUSCO	Suficiente	1,8	1,0
	Básico	17,4	5,1
	Previo	7,3	1,3
	Debajo de Previo	73,5	6,9
HUANCAVELICA	Suficiente	4,1	2,4
	Básico	19,8	6,4
	Previo	14,0	2,6
	Debajo de Previo	62,1	9,1
HUÁNUCO	Suficiente	1,6	1,1
	Básico	24,2	8,7
	Previo	6,5	1,4
	Debajo de Previo	67,8	9,4
ICA	Suficiente	10,3	3,3
	Básico	40,7	5,0
	Previo	14,2	2,9
	Debajo de Previo	34,9	6,5
JUNÍN	Suficiente	7,1	2,8
	Básico	42,0	5,0
	Previo	15,0	3,6
	Debajo de Previo	36,0	6,6

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	9,9	2,7
	Básico	39,2	4,1
	Previo	15,4	2,1
	Debajo de Previo	35,6	6,1
LAMBAYEQUE	Suficiente	8,9	0,9
	Básico	37,6	1,4
	Previo	14,2	0,7
	Debajo de Previo	39,3	1,7
LIMA	Suficiente	15,0	1,4
	Básico	49,6	2,1
	Previo	13,2	0,9
	Debajo de Previo	22,3	2,4
LORETO	Suficiente	0,4	0,4
	Básico	9,8	2,3
	Previo	9,7	2,4
	Debajo de Previo	80,1	4,0
MADRE DE DIOS	Suficiente	3,0	1,7
	Básico	26,1	4,0
	Previo	19,3	4,7
	Debajo de Previo	51,6	6,1
MOQUEGUA	Suficiente	13,1	3,5
	Básico	53,7	2,8
	Previo	13,3	2,0
	Debajo de Previo	19,9	3,7
PASCO	Suficiente	9,1	3,0
	Básico	28,8	6,6
	Previo	13,3	2,3
	Debajo de Previo	48,8	9,2
PIURA	Suficiente	4,9	1,6
	Básico	34,0	3,8
	Previo	13,5	1,8
	Debajo de Previo	47,6	5,2
PUNO	Suficiente	4,4	2,7
	Básico	17,5	4,3
	Previo	11,1	2,6
	Debajo de Previo	67,0	8,0
SAN MARTÍN	Suficiente	3,5	1,9
	Básico	20,4	7,7
	Previo	11,0	2,2
	Debajo de Previo	65,1	10,1
TACNA	Suficiente	15,1	3,0
	Básico	53,2	3,7
	Previo	11,3	1,8
	Debajo de Previo	20,5	4,6
TUMBES	Suficiente	2,9	1,1
	Básico	33,4	5,5
	Previo	16,1	2,9
	Debajo de Previo	47,5	6,7
UCAYALI	Suficiente	0,8	0,5
	Básico	12,3	2,7
	Previo	11,7	2,5
	Debajo de Previo	75,2	3,8

Tercer grado de secundaria

COMUNICACIÓN INTEGRAL

Cuadro 5.27: Porcentaje de estudiantes por nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	15,1	0,65
Básico	23,8	0,65
Previo	26,9	0,60
< Previo	34,3	0,89

Cuadro 5.28: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente	13,4	0,74	16,7	0,88	-2,87	0,004
Básico	22,5	0,91	25,0	0,74	-2,13	0,033
Previo	26,8	0,80	26,9	0,77	-0,09	0,928
< Previo	37,3	1,08	31,4	1,04	3,94	0,000

Cuadro 5.29: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	10,1	0,66	34,1	1,77	-12,70	0,000
	21,7	0,77	31,7	1,07	-7,59	0,000
	28,5	0,70	20,7	1,03	6,26	0,000
	39.7	1.03	13.5	1.6	13.77	0.000

Cuadro 5.30: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	11,1	0,75	3,3	0,55	8,39	0,000
	23,4	0,87	9,2	0,83	11,81	0,000
	29,4	0,78	21,6	1,3	5,14	0,000
	36,1	1,13	65,9	2,1	-12,50	0,000

Cuadro 5.31: Porcentaje de estudiantes por departamento según nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	9,2	2,1
	Básico	16,3	2,6
	Previo	23,8	2,7
	Debajo de Previo	50,7	5,8
ÁNCASH	Suficiente	9,5	2,2
	Básico	20,3	2,8
	Previo	26,0	2,6
	Debajo de Previo	44,2	5,6
APURÍMAC	Suficiente	3,8	1,4
	Básico	14,2	3,1
	Previo	20,8	1,6
	Debajo de Previo	61,3	5,0
AREQUIPA	Suficiente	22,2	3,4
	Básico	33,3	2,3
	Previo	26,2	2,5
	Debajo de Previo	18,4	2,6
AYACUCHO	Suficiente	9,5	3,4
	Básico	18,5	3,4
	Previo	25,9	2,5
	Debajo de Previo	46,1	6,8
CAJAMARCA	Suficiente	7,7	2,4
	Básico	16,7	2,9
	Previo	24,8	2,1
	Debajo de Previo	50,8	4,6
CALLAO	Suficiente	22,4	4,2
	Básico	31,0	1,6
	Previo	32,3	2,8
	Debajo de Previo	14,2	2,0
CUSCO	Suficiente	11,7	3,6
	Básico	20,7	2,5
	Previo	24,0	1,9
	Debajo de Previo	43,7	5,5
HUANCAVELICA	Suficiente	5,6	2,3
	Básico	15,3	4,6
	Previo	16,1	1,8
	Debajo de Previo	63,0	6,5
HUÁNUCO	Suficiente	8,3	3,5
	Básico	16,4	3,5
	Previo	23,3	2,7
	Debajo de Previo	52,0	6,8
ICA	Suficiente	8,2	1,1
	Básico	23,1	2,7
	Previo	33,4	1,8
	Debajo de Previo	35,3	2,8
JUNÍN	Suficiente	12,6	2,7
	Básico	24,3	2,4
	Previo	29,7	2,2
	Debajo de Previo	33,4	4,1

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	18,8	3,3
	Básico	27,9	2,1
	Previo	28,9	2,3
	Debajo de Previo	24,4	3,5
LAMBAYEQUE	Suficiente	13,3	2,9
	Básico	25,6	2,4
	Previo	28,6	2,5
	Debajo de Previo	32,5	4,0
LIMA	Suficiente	21,4	1,8
	Básico	28,6	1,8
	Previo	29,6	1,7
	Debajo de Previo	20,5	2,0
LORETO	Suficiente	6,4	1,9
	Básico	17,7	2,8
	Previo	27,8	2,2
	Debajo de Previo	48,1	4,3
MADRE DE DIOS	Suficiente	16,5	4,3
	Básico	22,9	6,5
	Previo	25,7	2,7
	Debajo de Previo	34,9	8,8
MOQUEGUA	Suficiente	26,2	6,7
	Básico	27,6	2,3
	Previo	26,0	3,3
	Debajo de Previo	20,2	5,8
PASCO	Suficiente	11,9	2,9
	Básico	21,5	3,6
	Previo	26,0	2,4
	Debajo de Previo	40,7	5,5
PIURA	Suficiente	13,9	2,5
	Básico	22,5	2,4
	Previo	24,0	2,1
	Debajo de Previo	39,6	4,7
PUNO	Suficiente	4,1	1,8
	Básico	10,2	2,0
	Previo	18,4	2,0
	Debajo de Previo	67,4	4,3
SAN MARTÍN	Suficiente	12,8	3,3
	Básico	20,2	2,9
	Previo	26,9	2,1
	Debajo de Previo	40,1	4,9
TACNA	Suficiente	29,7	4,7
	Básico	31,4	3,2
	Previo	24,1	2,4
	Debajo de Previo	14,8	3,3
TUMBES	Suficiente	13,8	5,6
	Básico	19,6	2,9
	Previo	29,7	2,5
	Debajo de Previo	37,0	6,9
UCAYALI	Suficiente	10,5	3,0
	Básico	20,1	2,6
	Previo	25,2	2,4
	Debajo de Previo	44,3	4,7

Cuadro 5.32: Porcentaje de estudiantes por nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	6,0	0,42
Básico	19,9	0,65
Previo	19,0	0,47
< Previo	55,1	1,02

Cuadro 5.33: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	6,7	0,54	5,2	0,52	2,00	0,045
	20,0	0,84	19,8	0,84	0,17	0,866
	19,4	0,63	18,6	0,64	0,89	0,373
	53.9	1.20	56.4	1.30	-1.41	0.158

Cuadro 5.34: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	2,2	0,31	20,7	1,60	-11,35	0,000
	15,1	0,72	38,3	1,46	-14,25	0,000
	19,2	0,53	18,2	0,99	0,89	0,373
	63,5	1,14	22,8	2,15	16,72	0,000

Cuadro 5.35: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	2,4	0,35	0,2	0,10	6,04	0,000
	16,5	0,81	5,2	0,79	9,99	0,000
	20,3	0,59	11,3	0,82	8,91	0,000
	60.8	1,28	83.3	1.32	-12.24	0.000

Cuadro 5.36: Porcentaje de estudiantes por departamento, dentro de cada nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	1,0	0,5
	Básico	12,8	2,7
	Previo	15,9	2,6
	Debajo de Previo	70,3	4,9
ÁNCASH	Suficiente	3,2	1,7
	Básico	13,2	2,2
	Previo	16,8	2,5
	Debajo de Previo	66,9	4,7
APURÍMAC	Suficiente	0,1	0,1
	Básico	8,3	2,8
	Previo	16,4	2,8
	Debajo de Previo	75,2	5,3
AREQUIPA	Suficiente	9,2	2,3
	Básico	33,8	3,9
	Previo	20,3	1,8
	Debajo de Previo	36,7	5,0
AYACUCHO	Suficiente	4,3	2,5
	Básico	14,0	4,6
	Previo	16,8	2,3
	Debajo de Previo	64,8	6,7
CAJAMARCA	Suficiente	1,5	1,2
	Básico	13,5	3,0
	Previo	18,3	2,0
	Debajo de Previo	66,7	4,9
CALLAO	Suficiente	8,2	2,9
	Básico	28,2	3,7
	Previo	23,9	2,2
	Debajo de Previo	39,7	4,8
CUSCO	Suficiente	3,6	1,7
	Básico	13,8	3,3
	Previo	17,6	2,4
	Debajo de Previo	65,0	6,0
HUANCAVELICA	Suficiente	5,3	4,1
	Básico	7,2	2,1
	Previo	15,1	3,4
	Debajo de Previo	72,5	6,5
HUÁNUCO	Suficiente	4,1	2,4
	Básico	11,3	4,0
	Previo	14,8	1,7
	Debajo de Previo	69,9	6,9
ICA	Suficiente	4,0	2,1
	Básico	14,5	2,5
	Previo	18,8	1,2
	Debajo de Previo	62,8	4,3
JUNÍN	Suficiente	8,5	3,0
	Básico	23,0	3,7
	Previo	21,4	2,4
	Debajo de Previo	47,0	5,9

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	5,6	1,6
	Básico	22,3	3,5
	Previo	24,9	2,3
	Debajo de Previo	47,2	5,2
LAMBAYEQUE	Suficiente	5,7	2,3
	Básico	16,9	3,5
	Previo	20,2	2,4
	Debajo de Previo	57,2	5,4
LIMA	Suficiente	9,1	1,2
	Básico	26,8	1,6
	Previo	21,8	1,1
	Debajo de Previo	42,4	2,5
LORETO	Suficiente	0,7	0,3
	Básico	10,4	3,6
	Previo	11,2	2,3
	Debajo de Previo	77,7	5,7
MADRE DE DIOS	Suficiente	3,9	3,7
	Básico	24,7	6,8
	Previo	17,9	2,1
	Debajo de Previo	53,5	8,8
MOQUEGUA	Suficiente	13,2	6,1
	Básico	25,4	5,3
	Previo	23,7	2,7
	Debajo de Previo	37,7	9,0
PASCO	Suficiente	7,2	3,4
	Básico	18,9	4,4
	Previo	22,4	2,7
	Debajo de Previo	51,5	7,8
PIURA	Suficiente	5,1	1,7
	Básico	18,8	3,3
	Previo	14,8	1,7
	Debajo de Previo	61,3	5,2
PUNO	Suficiente	1,9	0,9
	Básico	7,0	2,0
	Previo	10,4	1,9
	Debajo de Previo	80,7	4,2
SAN MARTÍN	Suficiente	2,3	1,0
	Básico	11,8	2,3
	Previo	19,7	2,7
	Debajo de Previo	66,2	4,9
TACNA	Suficiente	13,8	4,2
	Básico	37,6	6,1
	Previo	18,9	2,5
	Debajo de Previo	29,7	7,3
TUMBES	Suficiente	4,5	3,8
	Básico	13,9	3,2
	Previo	15,5	1,1
	Debajo de Previo	66,2	7,1
UCAYALI	Suficiente	1,6	1,4
	Básico	9,1	3,0
	Previo	14,5	2,4
	Debajo de Previo	74,8	6,0

Quinto grado de secundaria

COMUNICACIÓN

Cuadro 5.37: Porcentaje de estudiantes por nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	9,8	0,46
Básico	45,3	0,73
Previo	14,8	0,44
< Previo	30,1	0,85

Cuadro 5.38: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente	8,5	0,50	11,0	0,65	-3,05	0,002
Básico	42,8	0,99	47,6	0,88	-3,62	0,000
Previo	15,6	0,57	14, 0	0,58	1,97	0,049
< Previo	33,1	1,01	27,4	1,08	3,85	0,000

Cuadro 5.39: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente	7,0	0,48	20,0	1,21	-9,99	0,000
Básico	42,0	0,87	57,3	1,14	-10,67	0,000
Previo	15,9	0,53	10,7	0,71	5,87	0,000
< Previo	35,0	1,03	12,0	1,12	15,12	0,000

Cuadro 5.40: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	IE Estatales					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	7,6	0,53	2,1	0,37	8,51	0,000
	44,1	0,95	24,0	1,51	11,27	0,000
	16,1	0,58	14,8	0,92	1,20	0,232
	32,2	1,12	59,1	2,12	-11,22	0,000

Cuadro 5.41: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	IE Estatales urbanas					
	Pobres		No pobres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	3,1	0,36	10,2	0,77	-8,35	0,000
	31,9	1,18	50,9	1,05	-12,03	0,000
	17,5	0,84	15,3	0,78	1,92	0,055
	47,5	1,36	23,6	1,19	13,23	0,000

**Cuadro 5. 42: Porcentaje de estudiantes por departamento
según nivel de desempeño**

Departamento	Nivel de desempeño	Porcentaje
AMAZONAS	Suficiente	0,3
	Básico	6,6
	Previo	16,7
	Debajo de Previo	76,5
ÁNCASH	Suficiente	3,2
	Básico	6,5
	Previo	13,4
	Debajo de Previo	77,0
APURÍMAC	Suficiente	1,4
	Básico	4,9
	Previo	15,7
	Debajo de Previo	78,0
AREQUIPA	Suficiente	3,4
	Básico	18,1
	Previo	23,1
	Debajo de Previo	55,4
AYACUCHO	Suficiente	3,0
	Básico	10,1
	Previo	15,8
	Debajo de Previo	71,2
CAJAMARCA	Suficiente	1,2
	Básico	8,1
	Previo	13,2
	Debajo de Previo	77,5
CALLAO	Suficiente	3,3
	Básico	16,4
	Previo	27,7
	Debajo de Previo	52,7
CUSCO	Suficiente	2,4
	Básico	9,7
	Previo	17,2
	Debajo de Previo	70,7
HUANCAVELICA	Suficiente	2,3
	Básico	4,7
	Previo	14,6
	Debajo de Previo	78,5
HUÁNUCO	Suficiente	1,9
	Básico	9,4
	Previo	11,7
	Debajo de Previo	77,1
ICA	Suficiente	0,8
	Básico	6,5
	Previo	14,5
	Debajo de Previo	78,2
JUNÍN	Suficiente	4,6
	Básico	15,1
	Previo	19,6
	Debajo de Previo	60,7

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	12,4	2,2
	Básico	50,4	2,8
	Previo	13,8	1,6
	Debajo de Previo	23,5	2,8
LAMBAYEQUE	Suficiente	11,6	2,7
	Básico	42,1	2,6
	Previo	14,9	1,6
	Debajo de Previo	31,5	3,7
LIMA	Suficiente	13,2	1,2
	Básico	53,2	1,7
	Previo	14,4	1,2
	Debajo de Previo	19,2	2,0
LORETO	Suficiente	5,5	2,0
	Básico	32,7	3,4
	Previo	19,2	2,1
	Debajo de Previo	42,7	4,4
MADRE DE DIOS	Suficiente	5,1	2,8
	Básico	52,0	4,9
	Previo	16,7	6,7
	Debajo de Previo	26,3	9,0
MOQUEGUA	Suficiente	17,7	4,0
	Básico	47,5	3,6
	Previo	15,0	1,7
	Debajo de Previo	19,8	4,7
PASCO	Suficiente	6,8	1,6
	Básico	46,4	5,0
	Previo	15,3	1,1
	Debajo de Previo	31,5	5,9
PIURA	Suficiente	8,1	1,4
	Básico	43,4	3,8
	Previo	16,8	2,2
	Debajo de Previo	31,8	4,7
PUNO	Suficiente	1,6	0,6
	Básico	20,1	3,5
	Previo	13,5	1,3
	Debajo de Previo	64,8	4,2
SAN MARTÍN	Suficiente	7,7	2,2
	Básico	37,8	3,4
	Previo	17,8	1,6
	Debajo de Previo	36,7	4,3
TACNA	Suficiente	21,3	4,2
	Básico	54,9	3,6
	Previo	10,2	2,1
	Debajo de Previo	13,6	3,6
TUMBES	Suficiente	6,9	3,2
	Básico	47,1	8,7
	Previo	15,0	1,7
	Debajo de Previo	31,0	7,1
UCAYALI	Suficiente	3,7	0,9
	Básico	38,7	3,7
	Previo	18,4	1,8
	Debajo de Previo	39,2	4,0

Cuadro 5.43: Porcentaje de estudiantes por nivel de desempeño a nivel nacional

Niveles de desempeño	Nacional	
	%	e.e.
Suficiente	2,9	0,27
Básico	11,0	0,53
Previo	17,7	0,62
< Previo	68,5	1,06

Cuadro 5.44: Porcentaje de estudiantes por sexo según nivel de desempeño

Niveles de desempeño	Nacional					
	Hombres		Mujeres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	3,5	0,44	2,4	0,27	2,13	0,033
	12,0	0,66	10,1	0,72	1,95	0,052
	17,8	0,67	17,5	0,87	0,27	0,785
	66.7	1.23	70.0	1.38	-1.79	0.074

Cuadro 5.45: Porcentaje de estudiantes por gestión según nivel de desempeño

Niveles de desempeño	Nacional					
	Estatales		No estatales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	0,8	0,15	10,5	1,13	-8,51	0,000
	6,4	0,51	28,3	1,55	-13,42	0,000
	15,6	0,74	25,2	1,05	-7,47	0,000
	77.2	1.17	36.0	2.26	16.19	0.000

Cuadro 5.46: Porcentaje de estudiantes por área geográfica según nivel de desempeño

Niveles de desempeño	Nacional					
	Urbanas		Rurales		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	0,9	0,17	0,0			
	6,9	0,56	1,8	0,35	7,72	0,000
	16,5	0,82	7,8	0,76	7,78	0,000
	75.7	1.30	90.4	0.93	-9.20	0.000

Cuadro 5.47: Porcentaje de estudiantes por grupo de pobreza según nivel de desempeño

Niveles de desempeño	IE Estatales urbanas					
	Pobres		No pobres		Z	p
	%	e.e.	%	e.e.		
Suficiente Básico Previo < Previo	0,1	0,09	1,3	0,42	-2,79	0,005
	2,5	0,35	9,3	0,65	-9,21	0,000
	11,8	0,78	19,2	1,14	-5,36	0,000
	85,6	0,85	70,2	1,76	7,88	0,000

Cuadro 5. 48: Porcentaje de estudiantes por departamento según nivel de desempeño

Departamento	Nivel de desempeño	Porcentaje	Error estándar
AMAZONAS	Suficiente	0,3	0,2
	Básico	6,6	1,4
	Previo	16,7	3,8
	Debajo de Previo	76,5	5,0
ÁNCASH	Suficiente	3,2	2,1
	Básico	6,5	2,4
	Previo	13,4	2,1
	Debajo de Previo	77,0	5,3
APURÍMAC	Suficiente	1,4	1,0
	Básico	4,9	2,6
	Previo	15,7	4,0
	Debajo de Previo	78,0	6,3
AREQUIPA	Suficiente	3,4	1,3
	Básico	18,1	3,1
	Previo	23,1	1,8
	Debajo de Previo	55,4	4,6
AYACUCHO	Suficiente	3,0	2,0
	Básico	10,1	4,2
	Previo	15,8	3,3
	Debajo de Previo	71,2	8,5
CAJAMARCA	Suficiente	1,2	0,6
	Básico	8,1	1,9
	Previo	13,2	2,7
	Debajo de Previo	77,5	4,3
CALLAO	Suficiente	3,3	1,8
	Básico	16,4	3,2
	Previo	27,7	2,0
	Debajo de Previo	52,7	5,2
CUSCO	Suficiente	2,4	0,9
	Básico	9,7	3,6
	Previo	17,2	2,0
	Debajo de Previo	70,7	5,7
HUANCAVELICA	Suficiente	2,3	1,7
	Básico	4,7	3,5
	Previo	14,6	3,0
	Debajo de Previo	78,5	6,7
HUÁNUCO	Suficiente	1,9	1,9
	Básico	9,4	4,2
	Previo	11,7	2,7
	Debajo de Previo	77,1	6,7
ICA	Suficiente	0,8	0,6
	Básico	6,5	2,0
	Previo	14,5	2,9
	Debajo de Previo	78,2	5,2
JUNÍN	Suficiente	4,6	1,5
	Básico	15,1	3,5
	Previo	19,6	2,7
	Debajo de Previo	60,7	5,9

Departamento	Nivel de desempeño	Porcentaje	Error estándar
LA LIBERTAD	Suficiente	2,3	1,0
	Básico	11,9	2,7
	Previo	20,8	2,6
	Debajo de Previo	65,0	5,2
LAMBAYEQUE	Suficiente	3,8	2,2
	Básico	9,4	2,7
	Previo	16,5	2,8
	Debajo de Previo	70,4	5,7
LIMA	Suficiente	4,2	0,7
	Básico	14,1	1,3
	Previo	19,4	1,7
	Debajo de Previo	62,4	2,9
LORETO	Suficiente	0,7	0,5
	Básico	4,2	1,5
	Previo	10,9	2,6
	Debajo de Previo	84,3	4,3
MADRE DE DIOS	Suficiente		
	Básico	7,7	5,4
	Previo	30,9	9,9
	Debajo de Previo	61,5	9,3
MOQUEGUA	Suficiente	3,8	1,2
	Básico	13,8	3,7
	Previo	28,3	3,6
	Debajo de Previo	54,2	7,6
PASCO	Suficiente	3,9	2,1
	Básico	13,8	4,8
	Previo	17,1	2,3
	Debajo de Previo	65,2	8,4
PIURA	Suficiente	2,6	0,8
	Básico	9,4	1,8
	Previo	15,2	2,1
	Debajo de Previo	72,8	4,1
PUNO	Suficiente	0,4	0,3
	Básico	2,8	0,9
	Previo	10,7	1,9
	Debajo de Previo	86,2	2,5
SAN MARTÍN	Suficiente		
	Básico	3,4	1,3
	Previo	16,2	3,7
	Debajo de Previo	80,4	4,5
TACNA	Suficiente	6,5	3,0
	Básico	22,6	5,4
	Previo	25,8	2,5
	Debajo de Previo	45,1	7,4
TUMBES	Suficiente	1,2	0,8
	Básico	8,5	3,5
	Previo	15,1	3,5
	Debajo de Previo	75,2	6,1
UCAYALI	Suficiente	0,9	0,9
	Básico	2,2	1,0
	Previo	10,2	2,0
	Debajo de Previo	86,8	3,5

Conclusiones



Al igual que las evaluaciones nacionales anteriores, los resultados de la EN 2004 muestran problemas importantes de calidad y de equidad en los logros de los estudiantes en Comprensión de textos escritos y Matemática en todos los grados evaluados: la mayoría de estudiantes no alcanza los niveles de desempeño esperados para el grado. Este problema afecta a estudiantes de todos los estratos estudiados: instituciones urbanas y rurales, estatales y no estatales, varones y mujeres. Si bien esta afirmación resulta preocupante, lo es aún más la constatación de que en casi todos los grados y áreas evaluadas la mayor parte de los estudiantes del país se encuentra bastante lejos de lograr un nivel adecuado de dominio de las capacidades evaluadas (más del 50% se ubica entre el nivel previo y el grupo por debajo del previo). En otras palabras, superar estos resultados constituye un enorme desafío para la política educativa de los próximos años, a un mayor que si la mayoría de los estudiantes se encontrara en el nivel básico.

Uno de los retos que debe enfrentar la política educativa en relación a la evidencia encontrada sobre el alto grado de variabilidad que presenta el rendimiento de los estudiantes en ambas áreas —que se traduce a nivel de las aulas— supone el desarrollar capacidades en los docentes que les permitan contar con las herramientas pertinentes para que, considerando dichas diferencias, puedan aprovecharlas con el fin de generar aprendizajes significativos en cada uno de sus estudiantes.

En términos de equidad, los estudiantes provenientes de niveles socioeconómicos más desfavorecidos también mostraron resultados más bajos en los aprendizajes evaluados. Al interior del sector estatal también se evidencia un problema de equidad, ya que los estudiantes de las instituciones educativas ubicadas en entornos rurales —que tienden a asistir a escuelas multigrado— obtienen resultados más bajos respecto de aquellos que estudian en instituciones ubicadas en zonas urbanas. Estos resultados sugieren que los factores socioeconómicos constituyen una dimensión restrictiva importante de los rendimientos escolares. En tal sentido, es imprescindible que el sistema educativo, a través de una política intersectorial efectiva —y no meramente formal, como lo es actualmente—, garantice que los programas educativos se ajusten a las características y necesidades de la población más desfavorecida con el fin de cumplir su objetivo de constituirse en un sistema de igualdad de oportunidades.

Los resultados de la EN 2004 han puesto en evidencia el bajo nivel de logro alcanzado por los estudiantes al inicio de la educación primaria. Esta constatación revela que los bajos niveles de aprendizaje encontrados en evaluaciones anteriores —nacionales e internacionales— de grados intermedios y finales de la educación básica no son sino la expresión acumulada de las dificultades que los niños y niñas presentan ya en los primeros grados de escolaridad. Ello supone que la política educativa debería prestar atención particular a este problema. Entre otras medidas, se considera que esta debería dedicar

recursos y esfuerzos: *i)* para hacer que la educación inicial sea universal para los niños de 4 y 5 años, en particular en las zonas de mayor pobreza, con el fin de asegurar que los niños de estos ámbitos logren el aprestamiento indispensable (en lo referente a su nivel cognitivo, socioemocional y motriz) para que inicien de manera adecuada sus estudios primarios; *ii)* promover la asignación de aquellos docentes con más experiencia y/o alguna especialización en la enseñanza de la lectoescritura a los primeros grados de educación primaria; *iii)* reforzar los aspectos metodológicos y didácticos del enfoque comunicativo textual en los programas de capacitación docente en tanto la evidencia sugiere que los docentes todavía no se han apropiado adecuadamente del mismo.

La información presentada respecto a la evolución del rendimiento estudiantil en los últimos años indica, en términos generales, que entre 1998 y 2004 no hay diferencias significativas entre los rendimientos promedio en el área de Comunicación y Matemática. En el caso del área de Matemática de quinto de secundaria, se encontró una disminución muy pequeña en el rendimiento de los estudiantes. Si bien intentar explicar este hecho resulta bastante complejo, es importante señalar algunos elementos que aporten a la discusión. Por un lado, no deja de llamar la atención el hecho de que no se registren cambios en el rendimiento estudiantil a pesar del proceso de modernización educativa que se impulsó en el marco de los programas de mejoramiento que se iniciaron a mediados de la década de 1990 con apoyo de cooperación financiera del Banco Mundial y del Banco Interamericano de Desarrollo. Sin embargo, se debe tener en consideración que los sistemas educativos cambian lentamente. Las reformas impulsadas necesitan tiempo, continuidad y coherencia para hacerse realidad; es decir, aun cuando una reforma cumpla con estas características, existe un desfase inevitable entre las decisiones políticas y su traducción y expresión en los cambios a nivel de las aulas. Una vez que la enseñanza ha cambiado, las repercusiones en el aprendizaje de los estudiantes también demoran en observarse.

De otro lado, la caída registrada en el caso del quinto de secundaria debe analizarse en el marco del desorden curricular que caracterizó a dicho nivel en los primeros años de este gobierno y que propiciamente ha llevado al establecimiento del Diseño Curricular Nacional en el año 2005. Además, el rendimiento de los estudiantes de quinto de secundaria expresa la influencia acumulada de once años de escolarización; en otras palabras, su rendimiento en las evaluaciones expresa no solo el dominio de las capacidades del grado en el que se encuentran sino su aprendizaje acumulado a largo de su paso por la escuela.

Si bien, como se ha mencionado a lo largo de este informe, las diferencias en el rendimiento de los estudiantes están fuertemente asociadas a sus características socio económicas, ello no supone que las variables escolares al interior de las escuelas y de las propias aulas no tengan un efecto también en estas diferencias. En este sentido, lo encontrado en la EN 2004 para los sectores más desfavorecidos sugiere que existen ciertos aspectos sobre los cuales la política pedagógica puede intervenir para mejorar los aprendizajes de los estudiantes pertenecientes a estos sectores. Entre ellos se puede mencionar la relación profesor-alumno, las oportunidades de aprendizaje y las habilidades lectoras y matemáticas de los docentes.

Si bien la información de estos factores se presentará con detalle en otros documentos,¹¹² consideramos importante señalar algunas cosas. Del análisis de la información de

112. Al respecto se puede consultar el documento sobre el modelo de factores asociados al rendimiento estudiantil de la Evaluación Nacional 2004 que se publicará próximamente.

la EN 2004, se ha encontrado que un porcentaje considerable de las capacidades establecidas en el currículo no son desarrolladas en el aula. Esto afecta las oportunidades de aprendizaje de los alumnos,¹¹³ si se considera, además, que las capacidades más desarrolladas suelen ser trabajadas de manera operativa y mediante tareas de baja demanda cognitiva. Se encontró que los estudiantes suelen tener mayores dificultades para resolver aquellas preguntas que se vinculan con los contenidos menos desarrollados en el aula. En tal sentido, es importante que la visión y compromiso a largo plazo por parte de la política educativa se centre en sostener el apoyo a las reformas basadas en la calidad y en la equidad de oportunidades curriculares y no solamente en la igualdad al acceso a la educación escolarizada.

De otro lado, también se ha evidenciado que los docentes de los alumnos evaluados muestran dificultades en el dominio de algunas habilidades lectoras y matemáticas. La mayoría de ellos tiene capacidad de resolver solo las tareas más sencillas, asociadas a procedimientos mecánicos y directos. Además, se ha encontrado que hay una asociación entre los aprendizajes que muestran los estudiantes en la EN 2004 y las habilidades de sus profesores: los alumnos de los profesores con mayores habilidades en Matemática y Comunicación tienden a obtener, en promedio, mejores resultados en las pruebas respectivas. Se debe precisar que si bien la información presentada constituye un indicio importante a tomar en cuenta, esta debería ser profundizada por medio de estudios específicos sobre este tema. Así, entonces, es necesario que la política de capacitación docente busque mejorar no solo el aspecto metodológico de la práctica docente, sino también el desarrollo de su manejo disciplinario, y la didáctica y enfoque de las disciplinas que enseñan.

Si bien en los últimos años el Ministerio de Educación ha hecho un gran esfuerzo por articular un Diseño Curricular Nacional de la Educación Básica Regular, es necesario que este trabajo se afiance y complemente con el establecimiento de un sistema de estándares de desempeño que permita que el sistema educativo cuente con expectativas claras sobre lo que deben aprender nuestros estudiantes en cada grado, ciclo y área. En efecto, tal como señala el documento base para la formulación de lineamientos de Política Pedagógica, «una tarea pendiente y urgente es la formulación de un sistema de estándares en el ámbito nacional que defina los niveles de logro que se pretenden alcanzar al término de la EBR, y establecer criterios de evaluación con objetividad» (MED 2005: 120).

La formulación de estándares permitiría contar con un sistema de instrumentos que esclarezcan los propósitos y las metas curriculares de la EBR y permita una eficiente rendición de cuentas. Este sistema debe formularse de modo que permita coordinar todos los elementos e instrumentos del sistema educativo mediante definiciones claras, específicas y ejemplificadas, de lo que se considera una educación de calidad. Un sistema de instrumentos de este tipo permitiría orientar los recursos educacionales y haría posible la alineación y consistencia de las políticas referidas a la producción de textos escolares, a los programas de formación y capacitación de docentes, de las evaluaciones de aula y de sistema,¹¹⁴ además de orientar la vigilancia de los aprendizajes de sus hijos por parte de

113. Este aspecto y su influencia en el rendimiento estudiantil ha sido ampliamente estudiado por otros investigadores. Al respecto se puede consultar: Cueto, Santiago et al (2003). *Oportunidades de Aprendizaje y rendimiento en matemáticas*. Lima: GRADE; Cueto, Santiago et al (2006). *Oportunidades de Aprendizaje y rendimiento en comunicación integral en estudiantes de tercer y cuarto grado de primaria en Lima y Ayacucho*. Lima: GRADE.

114. Actualmente existe una enorme brecha entre las evaluaciones de aula que realizan los docentes y las evaluaciones a escala de sistema que se han presentado en el presente informe. Así, por ejemplo,

los padres de familia. Asimismo, el proceso de descentralización en marcha exige que el establecimiento de un sistema de estándares de desempeño nacional se inicie lo antes posible con el fin de contar con una base común nacional de las expectativas de logro de aprendizajes que permita orientar de manera más precisa los procesos de diversificación curricular a nivel regional y local. Las adaptaciones de las expectativas según las realidades locales deben hacerse para facilitar los propósitos académicos en distintos contextos nacionales, nunca deben consistir en disminuir o postergar las expectativas para subgrupos nacionales.

mientras la tasa de aprobación (evaluación de aula) promedio de sexto grado de primaria es de 91,8%, el porcentaje de estudiantes que se encuentra en el nivel suficiente (evaluación a escala de sistema) en Matemática es de 12,1%, en el caso de Comprensión de textos escritos, y de 7,9% en Matemática.

Bibliografía

- ADAMS, R. y M. Wu (editores)
2002 *PISA 2000 Technical Report*. OECD.
- ASMAD, U., D. PALOMINO, M. TAM y G. ZAMBRANO
2004 *Una aproximación a la alfabetización matemática y científica de los estudiantes peruanos de 15 años. Resultados del Perú en la evaluación internacional PISA. Documento de trabajo 10*. Lima: UMC - Ministerio de Educación
- BARRETT, P.
2001 *Assessing the reliability of rating data*. Consulta hecha en 25/04/2005. <<http://www.liv.ac.uk/~pbarrett/rater.pdf>>.
- BENAVIDES, M.
2002 «Para explicar las diferencias en el rendimiento en matemática de cuarto grado en el Perú urbano: Análisis de resultados partir de un modelo básico». En: J. Rodríguez y S. Vargas. (Eds.) *Análisis de los resultados y metodología de las pruebas CRECER 1998*, pp. 83-108. Lima. UMC- Ministerio de Educación.
- BOND, T. y CH. FOX
2001 *Applying the Rasch Model: Fundamental measurement in the Human Sciences*. Nueva Jersey: Lawrence Earlbaum Associates.
- BOUCHEY, H. y S. HARTER
2005 «Reflected Appraisals, Academic Self-Perceptions, and Math/Science Performance During Early Adolescence». *Journal of Educational Psychology*, vol.97, nº4 pp. 673-686.
- BURGA, A.
2005 *Evaluación del rendimiento estudiantil. Introducción a la teoría de Respuesta al Ítem*. Disponible en: <http://www.minedu.gob.pe/umc/articulosUMC.php>.
- CAIRNS, P.
2003 *MSC in Research Methods Statistics: Examples of Correlations*. UCL Interaction Centre. Consulta hecha en 26/04/2005. <<http://www.ucl.ac.uk/paul/PsyStats/4NonParaCorrel/4Examples.pdf>>.
- CARVAJAL, P., A. TREJOS, y J. SOTO
2004 «Búsqueda de la Relación Entre Áreas ICES en Matemáticas, Física, Lenguaje y Rendimiento en Matemáticas I y Matemáticas II a través del Análisis de Componentes Principales». *Scientia et Technica*, n.º 26, pp. 133 - 138.
- CHIEW, S. y B. FRASER
2000 «Teacher interpersonal behavior and elementary students' outcomes». *Journal of Research in Childhood Education*, vol.14, n.º2, pp. 216-231.

- COHEN, A. S. y S.H. KIM
1998 «An investigation of linking methods under the Graded Response Model». *Applied psychological Measurement*, vol.22 n.º2, pp.116-130.
- CONSEJO NACIONAL DE EDUCACIÓN
2005 *Hacia un Proyecto Educativo Nacional. Propuesta del Consejo Nacional de Educación*. Lima: CNE.
- CUESTA, M.
1996 «Unidimensionalidad». En J. Muñoz (ed.) *Psicometría*. Madrid: Pirámide, pp.239 - 291.
- CUETO, S,
2004 *Factores predictivos del rendimiento escolar, deserción e ingreso a educación secundaria en una muestra de estudiantes de zonas rurales del Perú*. Disponible en <http://epaa.asu.edu/epaa/v12n35/>
- EMBRETSON, S. y P. REISE
2000 *Item Response Theory for psychologists*. New Jersey: Lawrence Earlbaum Associates.
- ESCURRA, M.
1998 «Cuantificación de la validez de contenido por criterio de jueces». *Revista de Psicología*. Pontificia Universidad Católica, vol.6.
- ESPINOSA, G. y TORREBLANCA, A.
2004 *Cómo rinden los estudiantes peruanos en Comunicación y Matemática: Resultados de la Evaluación Nacional 2001. Informe descriptivo*. Lima: UMC - MED.
- FANTUZZO, J., G. DAVIS y M. GINSBURG
1995 ««Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement». *Journal of Educational Psychology*, vol.87, pp. 272-289.
- FILLMORE, L. y C. SNOW
2000 *What teachers need to know about language?* Washington: ERIC Clearinghouse on Languages and Linguistics.
- FISHER, W.
1992 «Reliability statistics». En J. Linacre (ed.) *Rasch Measurement Transactions Part 2*, 1996. Chicago: MESA Press, p.238.
- FOX, C.
1999 «An introduction to the partial credit model for developing nursing assessments». *Journal of Nursing Education*, vol.38, nº8, pp. 340-346.
- GONZÁLEZ, E. J.
2002 «Sistemas de Evaluación en América Latina». Instituto Nacional de Estudos e Pesquisas Educacionais— INEP. Consulta hecha en 24/05/2005. <www.inep.gov.br/download/internacional/idioma/Sistemas%20de%20Evaluación%20en%20américa%20latina.pdf>.
- GOOD, T. y J. BROPHY
1997 *Psicología educativa contemporánea*. 5.a ed. México: McGraw-Hill.
- GUADALUPE, C. (ed.)
2002 *La educación peruana a inicios del nuevo siglo. Documento de trabajo n° 12*. Lima: MECEP – MED.

- HAMBLETON, R., H. SWAMINATHAN y J. ROGERS
1991 *Fundamentals of Item Response Theory*. California: SAGE.
- HANSON, B.A. y A.A. BÉGUIN
2002 «Obtaining a common scale for Item Response Theory item parameters using separate versus concurrent estimation in the common-item equating design». *Applied Psychological Measurement*, vol.26 n.º1, pp.3-34.
- HERRERA, J.
(2002) *La pobreza en el Perú 2001. Una visión departamental*. Lima: Instituto Nacional de Estadística e Informática – Institut de recherche pour le développement.
- HILL, H., B. ROWAN y D. LOEWENBERG
2005 «Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement». *American Educational Research Journal*, vol.42, n.º. 2, pp. 371-406.
- INGEBO, G.
1997 *Probability in the measure of achievement*. Chicago: MESA.
- JARAMILLO, M. e ARTEAGA, I.
2003 «La inversión pública en educación: proceso de asignación y determinantes de la distribución del gasto por alumno». Informe final. Manuscrito. Lima: Instituto Apoyo.
- KISH, L.
1965 *Survey Sampling*. Nueva York: John Wiley & Sons.
- KOLEN, M.J. y R.L. BRENNAN
2004 *Test Equating, Scaling and Linking*. 2ª ed. Nueva York: Springer.
- LEGENDRE, P.
2005 «Species Associations: The Kendall Coefficient of Concordance Revisited». *Journal of Agricultural, Biological, and Environmental Statistics*, vol.10 n.º.2, pp.226–245.
- LINACRE, J.
2005 *Winsteps versión 3.57.2* [programa informático]. Chicago: Autor.
2000 «Comparing Partial Credit and Rating Scale Models». *Rasch Measurement Transactions*, vol.14 n.º 3, p.768. Consulta hecha en 03/07/2005. <<http://www.rasch.org/rmt/rmt143k.htm>>.
1994 «DIMTEST disminuyendo». *Rasch Measurement Transactions*, vol.8 n.º.3, p.384. Consulta hecha en 27/01/2005. <<http://www.rasch.org/rmt/rmt83n.htm>>.
- LINACRE, J. y Wright, B.
1994 «Chi-Square Fit Statistics». En J. Linacre (ed.) *Rasch Measurement Transactions Part 2*, 1996. Chicago: MESA Press, pp.360-361.
1989 «Length of a Logit». En J. Linacre (ed.) *Rasch Measurement Transactions Part 1*, 1995. Chicago: MESA Press, pp.54-55.
- LÓPEZ DE ULLIBARRI, I. y S. PITA
1996 «Medidas de concordancia: el Índice Kappa». *Cad Aten Primaria*, vol. 6, pp. 169-171.
- MACHINEA, J., A. BÁRCENA y A. LEÓN. Coordinadores.
2005 *Objetivos de Desarrollo del Milenio: una mirada desde América Latina y el Caribe*. En: <http://www.cepal.org/cgi-bin/getProd.asp?xml=/publicaciones/xml/1/21541/P21541.xml&xsl=/tpl/p9f.xsl&base=/tpl/top-bottom.xsl>

- MACMILLAN, P. D.
2000 «Classical, generalizability, and multifaceted rasch detection of interrater variability in large, sparse sets». *The Journal of Experimental Education*, vol.68 n.º2, pp.167-190.
- MARSH, H.
1990 *The Self Description Questionnaire II*. San Antonio, TX: Psychological Corporation.
- MARTINA, y M. FEMIA
2004 «Delta: A new measure of agreement between two raters». *British Journal of Mathematical & Statistical Psychology*, vol.57, pp.1-19.
- MARTIN, M.O, K.D. GREGORY y S.E. STEMLER
2000 *TIMSS 1999 Technical Report*. Boston: International Study Center.
- MINISTERIO DE EDUCACIÓN (MED)
2005a *Aproximaciones para un balance de la Educación Básica Regular en el Perú 1990 – 2004. Documento para la formulación de lineamientos de Política Pedagógica*. Lima: Comisión de Política Pedagógica – MED.
- MINISTERIO DE EDUCACIÓN – UNIDAD DE ESTADÍSTICA EDUCATIVA
2005b *Indicadores de la educación. Perú 2004*. Documento de Trabajo UEE n° 1 Lima: MED.
- MINISTERIO DE EDUCACIÓN – SECRETARÍA DE PLANIFICACIÓN ESTRATÉGICA
2002 *El Panorama Educativo*. Mimeo.
- MINISTERIO DE EDUCACIÓN (MED)
2001 *El desarrollo de la educación*. Informe nacional de la república del Perú elaborado por el Ministerio de Educación para la Oficina Internacional de Educación de UNESCO. Lima: MED.
- MONTERO, C. (coordinadora)
2001 *La escuela rural: Modalidades y prioridades de intervención*. Lima: MECEP – Ministerio de Educación.
- MUÑIZ, J.
1996 *Teoría Clásica de los Tests*, 2.a ed. Madrid: Ediciones Pirámide.
- NAVAS, M.J.
1996 «Equiparación de puntuaciones». En: J. Muñiz (coordinador) *Psicometría*, pp 293 – 370. Madrid: Universitas.
- NUNNALLY, J. y I. BERNSTEIN
1995 *Teoría Psicométrica*. 3.a ed. México: McGraw - Hill.
- OECD
2003 *Aptitudes básicas para el mundo del mañana. Otros resultados del proyecto PISA 2000. Resumen Ejecutivo*.
- OECD - SANTILLANA
2004 *Informe PISA 2003. Aprender para el mundo de mañana*. España: Santillana Educación.
- OREALC - UNESCO
2004 *La conclusión universal de la educación primaria en América Latina: ¿estamos realmente tan cerca?*. Santiago: OREALC.

- PATZ, R.J., B.W. JUNKER, M.S. JOHNSON y L. MARIANO
2002 «The hierarchical rater model for rated tests items and its application to large-scale educational assessment data». *Journal of Educational and Behavioral Statistics*, vol.27 n.º4, pp.341-384.
- PIZARRO, R., L. CLARK y M. ALLEN
1987 «El ambiente educativo del hogar». *Diálogos Educativos*, n.º 9-10, pp. 66-83.
- PRIETO, G. y A. GARCÍA
1996 «Construcción de Ítemes». En J. Muñiz (ed.) *Psicometría*. Madrid: Pirámide.
- SAAVEDRA, J. y SUÁREZ, P.
2002 *El financiamiento de la educación pública en el Perú: el rol de las familias*. Documento de trabajo n.º 38. Lima: GRADE.
- SAURA, P.
1995 *La educación del autoconcepto: cuestiones y propuestas*. Murcia: Universidad de Murcia.
- SCHULZ, E.
1990 «Functional assessment of fit». En J. Linacre (ed.) *Rasch Measurement Transactions Part 1*, 1995. Chicago: MESA Press, pp.82-84.
- SHAW, F.
1991 «Descriptive IRT vs. Prescriptive Rasch». En J. Linacre (ed.) *Rasch Measurement Transactions Part 1*, 1995. Chicago: MESA Press, p.131.
- STEEN, L. (ed)
1998 *La enseñanza agradable de la Matemática*. México, D. F.:MSEB / Limusa.
- STEMLER, S. E.
2004 «A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability». *Practical Assessment, Research & Evaluation*, vol.9, n.º4. Consulta hecha en 06/03/2005. <<http://PAREonline.net/getvn.asp?v=9&n=4>>
- THORNDIKE, R.
1989 *Psicometría aplicada*. México: Limusa.
- TORRES, R. M.
1995 *Repetición escolar: ¿falla del alumno o falla del sistema?* En: <http://www.mec.es/cide/rieme/documentos/torres1995re.pdf>.
- UNIDAD DE MEDICIÓN DE LA CALIDAD EDUCATIVA (UMC)
2005a *Evaluación Nacional del Rendimiento Estudiantil 2004. Informe Pedagógico de Resultados. Comprensión de textos escritos. Segundo grado de Primaria. Sexto grado de Primaria*. Lima: MED.
2005b *Evaluación Nacional del Rendimiento Estudiantil 2004. Informe Pedagógico de Resultados. Comprensión de textos escritos. Tercer grado de Secundaria. Quinto grado de Secundaria*. Lima:MED.
2005c *Evaluación Nacional del Rendimiento Estudiantil 2004. Informe Pedagógico de Resultados. Formación matemática. Segundo grado de Primaria. Sexto grado de Primaria*. Lima:MED.
2005d *Evaluación Nacional del Rendimiento Estudiantil 2004. Informe Pedagógico de Resultados. Formación matemática. Tercer grado de Secundaria. Quinto grado de Secundaria*. Lima: MED.

- 2005e *Evaluación Nacional del Rendimiento Estudiantil 2004. Una primera aproximación a la evaluación de la reflexión ciudadana. Sexto grado de Primaria.* Lima: MED.
- 2005f *Evaluación Nacional del Rendimiento Estudiantil 2004. Una primera aproximación a la evaluación de la reflexión ciudadana. Quinto grado de Secundaria.* Lima: MED.
- 2004 *Factores Asociados al Rendimiento Estudiantil. Resultados de la Evaluación Nacional 2001.* Lima: UMC - MED.
- VERHELST, N. y H. VERSTRALEN
- 1997 «Modeling Sums of Binary Responses by the Partial Credit Model». CITO. Consulta hecha en 21/02/2004. <<http://download.citogroep.nl/pub/pok/reports/Report97-7.pdf>>
- WATKINS, M. W.
2002. *MacKappa* [programa informático]. Pennsylvania State University: Autor.
- WAXMAN, H., S. HUANG, L. ANDERSON, T. WEINSTEIN
- 1999 «Classroom process differences in inner-city elementary schools». *The Journal of Educational Research*, vol. 91, n.º1, pp. 49-59.
- WENDLER C., L. YU, K. ZELLER y J. CAREY
- 2004 «Differential effects of the order of testing between high ability and low ability students». Presentado en la conferencia Annual de la American Educational Research Association, San Diego.
- WESTAT
- 2002 *WesVar versión 4.2* [programa informático]. Rockville: Autor
- WOLFE, E. W.
- 2004 «Identifying rater effects using latent trait models». *Psychology Science*, vol.46 n.º1, pp.35-51.
- WRIGHT, B.
- 1999 «Model selection: Rating Scale or Partial Credit?». *Rasch Measurement Transactions*, vol.12, n.º 3, p. 641-642. Consulta hecha en 03/07/2005. <<http://www.rasch.org/rmt/rmt1231.htm>>.
- 1989 «Useful Measurement through One-Step Item Banking.» En J. Linacre (ed.) *Rasch Measurement Transactions Part 1*, 1995. Chicago: MESA Press, p.24.
- WRIGHT, B. y J. LINACRE
- 1998 «MESA research memorandum 44». *Archives of Physical Medicine and Rehabilitation*, vol.70 n.º12, pp. 857-860.
- 1989 «The Differences between scores and measures». En J. Linacre (ed.) *Rasch Measurement Transactions Part 1*, 1995. Chicago: MESA Press, pp.63-65.
- WU, M., ADAMS, R. y WILSON, M.
- 1998 *ConQuest* [programa informático]. Melbourne: ACER: Autores
- YU, C.H. y S.E. ORBORN
- 2005 «Test equating by common items and common subjects: concepts and applications». *PARE*, vol.10 n.º4. Consulta hecha en 25/05/2005. <<http://pareonline.net/pdf/v10n4.pdf>>.

