

CNE OPINA

EVALUACIONES ESTANDARIZADAS DEL RENDIMIENTO ESCOLAR



El Consejo Nacional de Educación agradece al estudio Niños del Milenio (www.ninosdelmilenio.org), ejecutado por GRADE (www.grade.org.pe), por su ayuda en la provisión de información relevante para el presente boletín. Las opiniones que aquí se incluyen son, sin embargo, de exclusiva autoría del CNE.

Fotografía de portada: La Unidad de Medición de la Calidad Educativa

Este documento puede reproducirse para difusión y debate siempre y cuando se mencione la fuente.

Hecho el depósito legal en la Biblioteca Nacional del Perú N.º 2007-12962

Impreso en Impresión Arte Perú
Jr. Arnaldo Márquez 1899- Jesús María • Teléfono: 2615621

Índice

Resumen ejecutivo	2
I ■ Justificación y fines de las evaluaciones estandarizadas	5
II ■ Breve recuento de las evaluaciones en el Perú	9
III ■ ¿Cuál ha sido el rendimiento de los estudiantes peruanos?	12
IV ■ Balance y recomendaciones	29
V ■ Referencias	34
VI ■ Anexos	36

Resumen ejecutivo

La implementación de evaluaciones estandarizadas para los estudiantes es una práctica que se viene extendiendo en el mundo desde la segunda mitad del siglo pasado, y busca orientar la toma de decisiones en materia de políticas educativas e impactar positivamente en la calidad de la educación. El Perú implementa evaluaciones estandarizadas desde mediados del decenio de 1990. En la actualidad, las autoridades y la opinión pública consideran los resultados de estas evaluaciones como una referencia importante de la calidad de la educación.

En el presente documento, el CNE reflexiona sobre la finalidad de las evaluaciones estandarizadas y cómo se han implementado en el país. Luego, se analiza el rendimiento de los estudiantes peruanos en las principales evaluaciones. Finalmente, se presenta un conjunto de recomendaciones orientadas a que las evaluaciones estandarizadas tengan un impacto significativo y positivo en la calidad y la equidad de nuestro sistema educativo.

La formulación e implementación de un sistema de evaluaciones estandarizadas que ayude a mejorar un sistema educativo requiere establecer claramente qué es lo que se desea lograr con las evaluaciones. Existen dos grandes modelos al respecto. El primero involucra a familiares, docentes, directores y funcionarios de distinto nivel, y es denominado *modelo de bajas consecuencias*. En este modelo, los resultados de las pruebas no tienen implicancias directas, sino que se espera que sean utilizados por los diversos actores educativos para tomar decisiones que lleven a mejorar los procesos de enseñanza y aprendizaje. Este es el modelo adoptado en el Perú, y el Consejo Nacional de Educación lo considera como el más adecuado para nuestra realidad, a pesar de que necesita desarrollos y mejoras que se detallarán en la sección de recomendaciones.

El segundo, denominado *modelo de altas consecuencias*, usa los resultados para tomar decisiones que afectan directamente a los involucrados en la evaluación. En algunos sistemas educativos, los docentes de los estudiantes que alcanzan altos desempeños reciben incentivos monetarios o de cualquier otra naturaleza. El Consejo Nacional de Educación considera que, en el contexto peruano, una evaluación de altas consecuencias sería contraproducente, pues podría llevar a premiar a los actores educativos no por su esfuerzo, sino simplemente por contar con condiciones más favorables, como sucede con los estudiantes de zonas urbanas en comparación con los de zonas rurales. Cabe precisar que llama la atención del CNE que la Ley de Reforma Magisterial (Artículo 24) establezca que la evaluación del desempeño docente incluya la evaluación del progreso de los alumnos. Esto contradice al enfoque de bajas consecuencias y lo señalado en el Marco de Buen Desempeño Docente. Este último define las competencias que debe desarrollar el docente, pero no consigna los resultados del aprendizaje de los estudiantes como elementos de la evaluación docente. Esta discrepancia debe corregirse siguiendo los lineamientos del Marco de Buen Desempeño Docente.

En el Perú, las evaluaciones estandarizadas empezaron a implementarse y se desarrollaron bajo un enfoque de normas entre los años 1996 y 1998. Así, los test permitían identificar cuál era el desempeño de los estudiantes participantes en relación unos con otros, pero no permitían saber si habían logrado determinados niveles de desempeño. A pesar de esta limitación, las evaluaciones determinaron que los estudiantes de zonas rurales de habla distinta del español se encontraban notablemente rezagados en rendimiento.

Una segunda fase de evaluaciones estandarizadas se desarrolló desde el enfoque de criterios a partir de la medición con evaluaciones muestrales en los años 2001 y 2004. Esto significa que el diseño de las pruebas partió de un análisis del currículo nacional, gracias al cual se establecieron niveles de desempeño. A diferencia del enfoque de normas, el enfoque de criterios permite identificar el nivel de logro de los estudiantes participantes. En esta fase se produjo también un gran volumen de trabajos de análisis de los factores asociados, relevantes para orientar las políticas educativas. Mientras que la segunda fase de evaluaciones se administró a muestras representativas, la tercera fase se caracteriza por ser censal, de ahí su denominación “Evaluación Censal de Estudiantes”, que ha permitido tener resultados comparables desde 2008 hasta 2012, y que continúa en implementación.

El rendimiento de los estudiantes en las evaluaciones estandarizadas, tanto nacionales como internacionales, tiene dos características muy marcadas. En primer lugar, es un rendimiento bajo, tanto en relación con los niveles de logro esperados, como en la comparación con otros países. Un porcentaje mayoritario de estudiantes no está logrando los aprendizajes establecidos en el currículo. Otra característica es que el logro de aprendizajes está distribuido de forma inequitativa, pues los rendimientos más bajos los obtienen los estudiantes que provienen de familias que viven en zonas rurales y hablan lenguas distintas del castellano.

En la serie 2008-2012 de la Evaluación Censal de Estudiantes, si bien se identifica una tendencia hacia la mejora, que actualmente se ha estancado, se siguen presentando las brechas por zona geográfica y por lengua materna. Los logros de aprendizaje entre regiones son variables, y sus diferencias se correlacionan con el gasto invertido por estudiante. Por otra parte, cabe señalar que el objetivo de realizar una evaluación censal fue el brindar información a padres, docentes y funcionarios con la finalidad de que esta se revierta en toma de decisiones orientada a la mejora de los aprendizajes, pero este uso es bastante escaso. Asimismo, no se ha producido información sobre los factores asociados al rendimiento.

El Consejo Nacional de Educación considera que la implementación de evaluaciones estandarizadas ha sido positiva, puesto que, primero, ha contribuido a crear una cultura de la evaluación y la rendición de cuentas; segundo, ha permitido identificar a los sectores de la población que están más marginados del acceso a una educación de calidad; y, finalmente, ha generado una importante base de conocimientos sobre los factores que influyen en el rendimiento académico. Sin embargo, aun existe un amplio espacio para la mejora, y por ello el CNE plantea las siguientes recomendaciones:

- 1) Es necesario formular y difundir entre la comunidad expectativas precisas y detalladas de los rendimientos que deben lograr los estudiantes. Todos los componentes del sistema curricular nacional, que actualmente se encuentra en construcción, deben estar alineados con estas expectativas. Adicionalmente, sugerimos la implementación de un sistema de monitoreo de los procesos pedagógicos en las aulas que retroalimente el currículo.
- 2) Para avanzar hacia un mayor y mejor uso de los resultados, se requiere formular lineamientos para que las instituciones educativas establezcan sus metas y estrategias que les permitan desarrollar en los estudiantes los aprendizajes fundamentales suficientes para responder a las exigencias de las evaluaciones internacionales.
- 3) Parte importante del plan de evaluaciones deberían constituirlo evaluaciones muestrales que, acompañadas de cuestionarios detallados, permitan recoger información sobre el contexto del estudiante, de modo que sea posible profundizar en el conocimientos de los

factores asociados al desempeño, su peso específico, y sus interrelaciones en el logro de aprendizajes.

- 4) Debe formularse un plan de difusión de la información generada por las evaluaciones que incluya el desarrollo de capacidades para su uso adecuado. Este plan debe incluir a directores, docentes, padres de familia, autoridades locales, regionales y del nivel central, y también a quienes elaboran textos escolares y otros materiales didácticos y propuestas pedagógicas. También debería estimularse la participación de investigadores a fin de generar más conocimiento que ilumine la toma de decisiones.
- 5) Es imperativo continuar y reforzar la evaluación de las poblaciones que no hablan español, que son las más marginadas en el logro de aprendizajes.
- 6) Hace falta elaborar y difundir documentos técnicos que expliquen los fundamentos y propósitos de las evaluaciones, el marco muestral, los procedimientos logísticos, las especificaciones de pruebas y los procedimientos para calcular puntajes en las pruebas. Tal información es crucial, por un lado, para fortalecer la legitimidad de las evaluaciones, pero, por otro, para investigadores y otros actores que quieran usar la información en análisis particulares.
- 7) El organismo que tiene a su cargo la evaluación de los aprendizajes y la difusión de los resultados de esta evaluación debe alcanzar más autonomía e independencia del Ministerio de Educación, y debe contar con una asignación presupuestal acorde con la importancia de su labor. El CNE considera que este debe ser un tema a discutirse y propone, por el momento, la creación de un comité técnico independiente, integrado por personalidades de indiscutible calidad técnica y profesional.

Finalmente, es preciso enfatizar que es de especial interés para el CNE que se evite el uso político de los resultados de las evaluaciones, y que no se esgrima los resultados para legitimar o criticar determinados programas. La interpretación de los resultados requiere la máxima cautela, pues los logros educativos, además de los factores explicativos propiamente educativos, se deben también a una serie de factores externos de orden socioeconómico, nutricional, de salud, etc.

JUSTIFICACIÓN Y FINES DE LAS EVALUACIONES ESTANDARIZADAS

Las evaluaciones estandarizadas del rendimiento escolar son, en general, iniciativas costosas que requieren del trabajo dedicado de profesionales altamente especializados durante largos períodos, desde la concepción inicial hasta la elaboración de reportes para diferentes actores educativos. Por esta razón, es legítimo preguntarse si se debería apoyar o incluso reforzar esta práctica en el Perú. En opinión del Consejo Nacional de Educación, la respuesta es afirmativa, siempre y cuando esta práctica se enmarque en un plan comprehensivo de evaluaciones, que abarque varios grados y áreas del currículo, y siempre que se cuente con estrategias para su disseminación y utilización. Si bien el CNE reconoce que es mucho lo que se ha avanzado, pensamos que también hay mucho por hacer. Este tema es retomado al final del presente documento, después de discutir brevemente los enfoques dominantes actualmente acerca de las evaluaciones.

De acuerdo con diversos autores (por ejemplo, Ravela *et al.*, 2008), lo más importante al planificar evaluaciones es tener un plan claro de su propósito y de los fines que se les piensa dar. Al respecto, hay una gran variedad de posibilidades. Un primer enfoque es el denominado de *bajas consecuencias*. En este enfoque se planifican las evaluaciones para conocer cuánto han aprendido los estudiantes en relación con algún referente (por ejemplo, el currículo nacional). Para esta labor, a menudo se han establecido previamente estándares o expectativas de aprendizaje que se usan para calcular el porcentaje de estudiantes que llegan a un nivel satisfactorio de rendimiento. El propósito de este enfoque es únicamente formativo: los actores que tuvieran acceso a los resultados de evaluaciones deberían utilizarlos para, por ejemplo, identificar dificultades y diseñar políticas, revisar los programas de capacitación docente o revisar los textos escolares existentes, por mencionar algunos posibles usos. La escasa utilización de los resultados es uno de los principales problemas de este enfoque, pues esta depende de la motivación de los actores y de su experiencia en lectura e interpretación de información que a menudo viene en formatos poco comunes entre educadores (por ejemplo, información estadística).

En el segundo enfoque denominado *enfoque de evaluaciones de altas consecuencias*, los resultados tienen implicancias directas para algunos actores. **Por ejemplo, se puede administrar una evaluación a los estudiantes que terminan la secundaria**, ya sea para decidir su graduación o para tener un criterio referencial que se pueda usar para el ingreso a la educación superior o al mundo laboral. Este enfoque se justifica por la necesidad de evaluar a todos los estudiantes por igual, eliminando las diferencias provenientes de las preferencias de cada docente por distintos criterios. Así, en este enfoque se espera que padres de familia y docentes conozcan con anticipación los niveles de rendimiento necesarios para, por ejemplo, graduarse de un nivel educativo, y que trabajen para lograrlos.

Las evaluaciones estandarizadas del rendimiento escolar, en general, iniciativas costosas que requieren del trabajo dedicado de profesionales altamente especializados durante largos períodos.

Este tipo de evaluaciones podría representar una ventaja para los estudiantes que, en diferentes partes de nuestro país, podrían ser evaluados y, sobre la base de sus puntajes, postular a una o varias universidades, y no como ocurre actualmente, que deben someterse a evaluaciones específicas a cada universidad a la que quieran ingresar.

Este es el caso en El Salvador, donde se administra la Prueba de Aprendizaje y Aptitudes para Egresados de la Educación Media (PAES)¹. Se trata de una prueba obligatoria que tiene un peso del 25% para la nota final en las áreas evaluadas (matemática, estudios sociales, lenguaje y literatura, y estudios naturales). A este tipo de evaluación se le hacen al menos dos críticas. La primera es que, dado que se hace imposible planificar estas evaluaciones en todas las áreas del currículo, en la práctica se priorizan unas cuantas consideradas “instrumentales” para el aprendizaje a lo largo de la vida (típicamente, comprensión de lectura, matemática y ciencias). Sin embargo, al hacerse esto, y enfatizarse la enseñanza de las áreas evaluadas, pierde fuerza la pretensión de que la educación sea una labor integral que involucre a todas las áreas del currículo. La segunda crítica es que este tipo de evaluación limita las posibilidades de trabajo o estudio de los estudiantes que viven en pobreza o pobreza extrema, ya que son ellos los que suelen obtener rendimientos más bajos. Si bien es cierto que todos los estudiantes deben aprender al nivel prescrito por el currículo, hay que considerar el impacto social de las evaluaciones con altas consecuencias.

Una variante del anterior modelo es la evaluación estandarizada de todos los estudiantes al finalizar la secundaria, pero no para definir si se deben graduar, sino solo como referente para el ingreso a la universidad. Por ejemplo, en Chile, se administra desde el 2003 la Prueba de Selección Universitaria (PSU), que busca medir los niveles de razonamiento de los estudiantes. La PSU incluye, de manera obligatoria, evaluaciones en lenguaje y comunicación, y matemática; además, incluye, como electivas, pruebas de historia y ciencias sociales, y ciencias (física, química y biología)². Algo similar se ha hecho en Brasil con la prueba ENEM (Examen Nacional de la Enseñanza Media)³. Estas prácticas parecen, en principio, menos problemáticas que la anterior, siempre y cuando el puntaje no sea el único criterio para la toma de decisiones (las notas del colegio, la motivación y perseverancia, o la habilidad para trabajar en grupos también pueden predecir el éxito en la educación superior o el trabajo). De hecho, este tipo de evaluaciones podría representar una ventaja para los estudiantes que, en diferentes partes de nuestro país, podrían ser evaluados y, sobre la base de sus puntajes, postular a una o varias universidades, y no como ocurre actualmente, que deben someterse a evaluaciones específicas a cada universidad a la que quieran ingresar. Sin embargo, hay que recordar que la PSU ha sido duramente criticada por académicos y estudiantes chilenos, pues finalmente los que consiguen puntajes más altos son los de familias con mayores ingresos⁴; así, la PSU legitima inequidades sociales que se manifiestan en los resultados de las pruebas de habilidades educativas.

Las evaluaciones también puede tener altas consecuencias si se emplean para otorgar incentivos a los docentes, como se ha hecho por décadas en Chile y México (Ferrer, 2006), o para cerrar instituciones educativas, como se ha hecho con la iniciativa legislativa *Ningún Niño se Queda Atrás* (*No Child*

1 Ver http://www.mined.gob.sv/jdownloads/PAES%202012/documento__informativo_-_paes_2012.pdf para una descripción de la PAES 2012.

2 La PSU es administrada desde la Universidad de Chile, para más información ver <http://www.demre.cl/psu.htm>.

3 Ver <http://portal.inep.gov.br/web/enem/sobre-o-enem>.

4 Ver, por ejemplo, nota periodística en <http://cnchile.com/noticia/2013/01/04/las-principales-criticas-de-los-especialistas-sobre-los-resultados-de-la-psu>

Left Behind) en los Estados Unidos⁵. La justificación de estos usos reside en la economía, o, al menos, en una visión de esta: los docentes, junto con los directores y el personal de la escuela, tienen que generar un producto (aprendizaje). Si lo logran individual o colectivamente, deben ser premiados; en caso contrario, deben pasar por procesos de capacitación o ser despedidos. Desde esta perspectiva, la realidad educativa funciona como un mercado: las distintas empresas (escuelas) ofrecen un servicio (educación) a los clientes (familias), que tienen la libertad de elegir entre distintas opciones. En muchas de estas propuestas las familias reciben una asignación por cada hijo (voucher) y pueden con ella acceder a la escuela que les brinde el servicio que consideren más conveniente. El supuesto tras esta perspectiva es que la competencia entre las distintas empresas (escuelas) por atraer a mayor cantidad de clientes (familias) las llevará a ofrecer servicios de mayor calidad, más innovadores y más atractivos. La investigación sobre este tema es variada y no concluyente respecto de si efectivamente la competencia entre escuelas genera mejoras en los aprendizajes o simplemente lleva a una mayor segmentación, por la cual los estudiantes de familias más educadas son las que procuran las escuelas con mayor prestigio (ver, por ejemplo, Hsieh y Urquiola, 2006).

En el CNE, sin embargo, consideramos que adoptar el enfoque de evaluación con altas consecuencias, sobre todo en la variante recién mencionada, sería riesgoso por una serie de consecuencias identificadas en la literatura. En primer lugar, es altamente probable que la evaluación únicamente de las áreas consideradas instrumentales lleve a un reduccionismo de la educación. Nosotros pensamos que la evaluación debería reflejar la amplitud de conocimientos y habilidades prevista en el currículo peruano. No menos importante es que, para poder identificar el efecto docente (o el valor agregado del docente y la escuela), se tendría que evaluar a los mismos estudiantes en años sucesivos; de otra forma, se podría estar atribuyendo al docente aprendizajes estudiantiles que son consecuencia más de las características de los estudiantes que del trabajo del docente o de su equipo docente en el aula. Debemos tener presente que este tipo de medición del efecto docente para todos los estudiantes tendría un costo enorme, y que existiría la posibilidad de que las evaluaciones, al ser tan masivas, se hicieran con base en instrumentos y procedimientos logísticos de baja calidad.

En segundo lugar, la competencia que se pretende generar entre escuelas podría ocurrir en alguna medida en zonas urbanas. En zonas rurales, en cambio, no habría competencia posible, pues a menudo hay una sola institución educativa por localidad. Adicionalmente, si se prepara información para que los padres elijan, seguramente serán los más educados los que harán mejor uso de esta; así, sin pretenderlo, este modelo basado en la provisión de información se podría convertir en una nueva fuente de inequidad.

En tercer lugar, la provisión de incentivos en función del rendimiento de estudiantes podría llevar a políticas de discriminación en el ingreso de estudiantes que se percibe que podrían “bajar el promedio”; también podrían generar el ausentamiento de los de menor rendimiento al momento de la evaluación. Controlar estos problemas es posible, pero requiere de sistemas de monitoreo que actualmente son débiles en el Perú. Finalmente, se debe recordar que la investigación educativa, en el Perú y otros países en desarrollo, ha mostrado repetidamente que la pobreza y la etnicidad (v. gr. pertenencia a un grupo hablante de lengua indígena) son dos de los principales factores asociados al rendimiento. Se podrían hacer comparaciones entre grupos de estudiantes y escuelas similares, como se ha hecho en Chile, **pero la complejidad técnica y los costos de implementar un enfoque de altas consecuencias de una manera sólida nos llevan a sugerir que debe ser evitado en el momento actual en el Perú**⁶.

⁵ Ver <http://www2.ed.gov/nclb/landing.jhtml> para más información.

⁶ Para una mayor discusión de los riesgos y posibilidades de usos de altas consecuencias en educación, sugerimos ver Hamilton, Stecher & Klein, 2002, y para una discusión de diferentes modelos de rendición de cuentas ver Anderson, 2005.

Vinculado a lo anterior, el CNE expresa su preocupación por el contenido del artículo 24 de la Ley de Reforma Magisterial N.º 29944, que dice: “La evaluación del desempeño tiene como finalidad comprobar el grado de desarrollo de las competencias y desempeños profesionales del profesor de aula, la institución educativa y la comunidad. La evaluación se basa en criterios de buen desempeño docente contenidos en las políticas de evaluación establecidas por el Ministerio de Educación, *lo que incluye necesariamente la evaluación del progreso de los alumnos*” (resaltado nuestro). Esto sugeriría que es necesario evaluar a los estudiantes usando un enfoque de altas consecuencias. Sin embargo, en el Reglamento de la Ley (DS N.º 004-2013-ED), la Octava Disposición Complementaria Transitoria dice: “La aplicación de los resultados de las evaluaciones de desempeño como requisito para la participación del profesor en concursos o la obtención de diversos beneficios, queda en suspenso en tanto no se generalicen dichas evaluaciones por parte del Minedu”. El mismo reglamento especifica que la evaluación del desempeño se debe realizar considerando el Marco de Buen Desempeño Docente⁷. En este documento se definen cuatro dominios, nueve competencias y cuarenta desempeños; sin embargo, ninguno de ellos se vincula con los aprendizajes de los estudiantes. Por lo tanto, existe una incoherencia entre la Ley y el Reglamento. A diferencia de la ley, este último sí es coherente con el Marco de Buen Desempeño Docente.

La posición del CNE es que se debería seguir priorizando un enfoque de evaluación de bajas consecuencias, como se infiere del Marco de Buen Desempeño Docente.

La posición del CNE es que se debería seguir priorizando un enfoque de evaluación de bajas consecuencias, como se infiere del Marco de Buen Desempeño Docente. Esta posición se basa en el supuesto de que, si los docentes e instituciones educativas cuentan en cada localidad con insumos relevantes (por ejemplo, textos, infraestructura y capacitación, además de información sobre evaluaciones) y han aprendido diferentes formas para fomentar aprendizajes de acuerdo con lo previsto en el currículo, es razonable esperar que desarrollen un trabajo pedagógico rico en las aulas, lo cual debería llevar a mejores resultados en las evaluaciones. Así, mientras que el enfoque de altas consecuencias suele priorizar la responsabilización de actores (por ejemplo, docentes) y los incentivos; nuestra propuesta, en cambio, se basa en priorizar insumos y procesos pedagógicos de alta calidad, con énfasis en las poblaciones que tradicionalmente han tenido más bajo rendimiento. Queda la crítica, sin embargo, de que el modelo de bajas consecuencias implementado en el Perú ha producido notables informes técnicos que han sido relativamente poco utilizados. Al respecto haremos algunas recomendaciones en la parte final del presente documento.

⁷ Documento elaborado y publicado por el Ministerio de Educación, ver <http://www.perueduca.pe/documents/60563/ce664fb7-a1dd-450d-a43d-bd8cd65b4736>

BREVE RECUENTO DE LAS EVALUACIONES EN EL PERÚ

La evaluación estandarizada en el Perú ha atravesado por tres períodos. Al primero podríamos denominarlo *de los inicios*, y comprende las evaluaciones realizadas por el Estado peruano en 1996 y 1998. Estas evaluaciones se hicieron con un enfoque de normas. En este enfoque, el objetivo es diseñar pruebas que permitan medir habilidades, pero buscando que los puntajes se distribuyan siguiendo una curva normal, es decir, con la mayor parte alrededor del promedio y cada vez menos observaciones a medida que uno se aleja de este. Las evaluaciones referidas a normas, si bien permiten distinguir diferencias de rendimiento entre estudiantes, no permiten clasificarlos de acuerdo con su nivel de dominio de los contenidos curriculares; se trata de un enfoque muy útil para medir algunas características psicológicas (por ejemplo inteligencia), pero no tanto para situaciones educativas. Los resultados en este enfoque se presentan en categorías relativas, por ejemplo, alto, medio alto, medio bajo y bajo (si se clasificara en cuatro grupos de 25% de estudiantes cada uno).

Las pruebas utilizadas en el Perú contaban en general con ítems bien diseñados, pero que, en conjunto, no habían sido pensados para reflejar cuánto habían aprendido los estudiantes de lo prescrito en el currículo (esto es común en un enfoque de normas). Sin embargo, estas pruebas permitieron identificar grupos de estudiantes con menor rendimiento: los relativamente pobres, rurales e indígenas. Este patrón de resultados se ha repetido en cuanta evaluación se ha realizado desde entonces. En esta época, se realizó, además, la primera evaluación llevada a cabo por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Llece) de la Unesco. Tal vez por la poca comprensión que en el Estado y la opinión pública se tenía de las evaluaciones y por el momento político que se vivía en la segunda mitad de los 90, los resultados de la evaluación nacional de 1996 y de la primera evaluación de la Unesco fueron inicialmente ocultados por las autoridades de turno. Luego, sin embargo, se publicaron los resultados (UMC y Grade, 2001). En esta primera etapa se elaboraron además boletines informativos que utilizaron ítems de las pruebas de 1998 y estudios de factores asociados al rendimiento. Se puede encontrar información adicional sobre estas evaluaciones y las siguientes en el portal de la Unidad de Medición de la Calidad Educativa (UMC) del Ministerio de Educación⁸. **En sus más de 15 años de funcionamiento, la UMC ha logrado establecer altos estándares de calidad, y con ello ha ganado legitimidad para su trabajo y el reconocimiento del CNE.**

Estas pruebas permitieron identificar grupos de estudiantes con menor rendimiento: los relativamente pobres, rurales e indígenas. Este patrón de resultados se ha repetido en cuanta evaluación se ha realizado desde entonces.

8 Ver <http://umc.minedu.gob.pe/>.

Se dijo que el Perú había salido último del mundo en la evaluación, cuando en realidad había salido último en una evaluación de 65 países, principalmente miembros de la OCDE (es decir, de países industrializados y con niveles muy bajos de pobreza y repitencia).

La segunda etapa de las evaluaciones nacionales abarca desde el año 2000 hasta el 2006, y se podría denominar de *cambio de modelo*. En esta etapa, se dejó el modelo de *normas* para adoptar uno de criterios. Las evaluaciones referidas a criterios buscan reflejar con precisión el dominio de conocimientos y habilidades que los estudiantes deberían dominar para luego identificar si efectivamente lo han logrado. Así, los resultados se pueden reportar en función a categorías como “no logrado”, “básico”, “satisfactorio” y “avanzado”. La meta en este caso sería lograr que todos los estudiantes estén en los dos grupos superiores.

En concordancia con este enfoque, el diseño de las pruebas en el Perú a partir del año 2000 partió de un análisis del currículo nacional. Este es el enfoque que en general siguen actualmente los sistemas de evaluación educativa nacional e internacional. Para poder contar con pruebas de este tipo, se desarrollaron fundamentaciones para las evaluaciones, es decir, se explicaron sus objetivos, se planificó el uso de los resultados y se desarrollaron especificaciones para las pruebas detallando sus características. Parte de este proceso fue definir los puntos de corte en las escalas de puntaje. La UMC desarrolló este trabajo siguiendo criterios aceptados internacionalmente, que fijan estos puntos de corte por acuerdo entre expertos y de acuerdo con lo que prescribe el currículo, las características de la prueba y el grado que cursan los estudiantes. En esta etapa, entre el 2001 y el 2004, se realizaron evaluaciones en primaria y secundaria en diferentes áreas del currículo. Las evaluaciones permitieron una serie de análisis pedagógicos, que fueron publicados por la UMC, junto con análisis de factores asociados al rendimiento.

En esta segunda etapa, se inició además la participación de Perú en las evaluaciones del Programa Internacional de Evaluación de los Estudiantes, conocido como PISA por sus siglas en inglés. Esta prueba, diseñada por la Organización para la Cooperación y el Desarrollo Económico (OCDE), busca evaluar las habilidades de los estudiantes de 15 años en la secundaria en lectura, matemática y ciencias. A pesar de lo valioso de los marcos conceptuales de evaluación y las especificaciones de las pruebas de PISA, que mostraban los resultados (más adelante se presentan algunos de ellos) y lo que se espera que los estudiantes de países industrializados pudieran aprender, a menudo en el Perú solo se destacó nuestra posición en el *ranking*, con una visión negativa y pesimista (sobre todo en la prensa), y no se exploró en profundidad qué se podría aprender para mejorar el sistema educativo; incluso, en algunos casos, se tergiversó información, como cuando se dijo que el Perú había salido último del mundo en la evaluación, cuando en realidad había salido último en una evaluación de 65 países, principalmente miembros de la OCDE (es decir, de países industrializados y con niveles muy bajos de pobreza y repitencia). No se trata de una situación exclusiva del Perú; Ravela (2006) ha documentado cómo muchos países de la región presentan sus resultados de una manera muy negativa y, a menudo, con interpretaciones que no tienen fundamento en los datos.

Finalmente, se podría decir que la tercera etapa de las evaluaciones estandarizadas en el Perú ha sido la de *las evaluaciones censales*, que abarca desde el 2006 hasta la actualidad. A mediados del año 2006, cuando

asumió el poder un nuevo gobierno, se decretó la realización de una evaluación censal de los estudiantes de segundo grado de primaria en comprensión de lectura en castellano y matemática, y de estudiantes en escuelas de educación intercultural bilingüe (EIB) en cuarto grado. La justificación fue que era necesario conocer los niveles de aprendizaje de todos los estudiantes en estas áreas hacia el inicio de su escolaridad y devolver a cada institución educativa sus resultados, para que diferentes actores pudieran actuar con base en esta información⁹. En la evaluación del 2006 hubo problemas para diseñar las pruebas y aplicarlas a un universo tan grande en pocos meses. Esto demuestra una vez más la importancia de conceder plazos razonables para procesos complejos como determinar los niveles de rendimiento de la población de un determinado grado en el Perú.

A pesar de su nombre, las evaluaciones censales de segundo grado excluyen a las instituciones educativas donde hay menos de cinco estudiantes en segundo grado de primaria. Esta exclusión nos resulta objetable, pues, si bien se trata de relativamente pocos estudiantes, son precisamente los que viven en zonas rurales, es decir de presumible menor rendimiento.

Usando las bases del Padrón 2012 y el Censo Escolar 2012¹⁰, estimamos a cuánto ascendería el número de instituciones educativas y alumnos no evaluados debido a este criterio. Como se observa en el cuadro 1, a nivel nacional, el 28% de instituciones educativas del país tiene menos de cinco estudiantes en segundo grado, aunque el porcentaje de estudiantes que estas instituciones educativas atienden es de alrededor del 4%. Podría argumentarse que se trata de un porcentaje bajo de estudiantes, pero consideramos que se debería revisar esta exclusión, pues las escuelas con menor estudiantado se encuentran casi siempre en zonas rurales, relativamente pobres. Cubrir las todas, en una verdadera evaluación censal, sería seguramente muy costoso; la alternativa sería evaluar una muestra de estas escuelas. Hacer esto seguramente no alteraría mucho los resultados, pero sentaría un precedente respecto de la inclusión educativa en el Perú. En cualquier caso, todas las instituciones educativas deberían recibir luego copias de los instrumentos de evaluación. Volveremos sobre este punto al final del documento.

Cuadro 1. Escuelas y alumnos no evaluados en segundo de primaria (2012)

	Escuelas		Alumnos	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Con 1 a 4 alumnos en segundo de primaria	9722	28 %	24473	4 %
Con 5 o más alumnos en segundo de primaria	24642	72 %	613545	96 %
Total	34364 ¹¹	100 %	638018	100 %

Fuente: Cálculos propios con base en el Padrón 2012 y el Censo Escolar 2012.

9 Ver marco de referencia de las evaluaciones en http://www2.minedu.gob.pe/umc/ece/Marco_de_Trabajo_ECE.pdf.

10 Bases disponibles en la página de Estadística de la Calidad Educativa

11 No se tomaron en cuenta las escuelas que no tenían alumnos en segundo de primaria.

¿CUÁL HA SIDO EL RENDIMIENTO DE LOS ESTUDIANTES PERUANOS?

En esta sección, presentamos brevemente un recuento de los resultados de los estudiantes peruanos en las evaluaciones nacionales e internacionales de rendimiento escolar. No reportamos los resultados de las dos primeras evaluaciones nacionales mencionadas arriba (1996 y 1998), dado que el enfoque de normas no permite definir cuántos estudiantes se ubicaban en un nivel aceptable de rendimiento. Tampoco se presentan los resultados de la primera evaluación del Llece, sino que nos concentraremos en las evaluaciones realizadas desde el 2000.

Evaluaciones nacionales de 2001 y 2004

En ambos casos se trató de evaluaciones muestrales, que permitían reportar resultados para el país y desagregaciones por algunos estratos (por ejemplo, público-privado, urbano-rural, hombres-mujeres y regiones). El cuadro 2 presenta el porcentaje de estudiantes que logró el nivel “suficiente” (aceptable) en comunicación y matemática. Desafortunadamente, los datos no son estrictamente comparables entre grados un mismo año o entre años en la misma área, pero de todos modos sugieren un bajo rendimiento de la mayoría de estudiantes, especialmente en matemática.

Cuadro 2. Porcentaje de estudiantes con nivel “suficiente” en las evaluaciones nacionales 2001 y 2004

Grados	Evaluación Nacional 2001		Evaluación Nacional 2004	
	Comunicación	Matemática	Comunicación	Matemática
2.º de primaria			15,1	9,6
4.º de primaria	34,8	26,6		
6.º de primaria	9,1	7,1	12,1	7,9
3.º de secundaria			15,1	6,0
4.º de secundaria	21,3	5,2		
5.º de secundaria			9,8	2,9

Fuente: Informes de la UMC. / Nota: Las celdas en blanco indican que no se evaluó el área correspondiente.

En estas evaluaciones, se incluyeron otras áreas de aprendizaje: expresión oral (2001), expresión escrita (2001 y 2004), solución de problemas en matemática usando material concreto (2001) y formación ciudadana (2004). Se usaron ítems de opción múltiple, pero también de respuesta corta y larga (por ejemplo, para escritura o reflexión ciudadana). Adicionalmente se recolectaron datos sobre las actitudes de los estudiantes hacia la matemática, lectura, y poblaciones y lenguas indígenas de todos los estudiantes (2001), y de comprensión de lectura en lengua indígena (2001 y 2004). Finalmente, se recolectaron datos sobre el estudiante y su familia, las escuelas, los docentes y el entorno en ambas evaluaciones. Con toda esta información, se elaboraron análisis de diverso tipo: primero,

informes descriptivos con los resultados y, luego, análisis de factores asociados¹². En resumen, el análisis y producción de reportes de factores asociados fue intenso, algo que desafortunadamente no se ha repetido con las evaluaciones censales.

Evaluaciones internacionales

Como se mencionó antes, el Perú ha participado en dos evaluaciones del Llece. La más reciente, administrada el 2006, se denominó *Segundo Estudio Regional Comparativo y Explicativo (Serce)*. En esta evaluación, participaron dieciséis países en comprensión de lectura y matemática y nueve en ciencias. Los resultados se presentan en el cuadro 3 de acuerdo con niveles de rendimiento (el nivel IV es el más alto; no hemos encontrado en la descripción del Llece un criterio para definir a partir de cuál se podría considerar un rendimiento aceptable).

Cuadro 3: Resultados de los estudiantes peruanos en el Segundo Estudio Regional Comparativo y Explicativo (2006)

Nivel	Matemática		Lectura		Ciencias Naturales
	3.º grado	6.º grado	3.º grado	6.º grado	6.º grado
Nivel IV	4,8 %	9,3 %	3,7 %	9,5 %	0,4 %
Nivel III	8,6 %	28,9 %	15,1 %	22,6 %	6,4 %
Nivel II	26,0 %	39,8 %	35,8 %	41,7 %	39,4 %
Nivel I	45,2 %	19,6 %	36,2 %	24,1 %	46,9 %
< Nivel I	15,4 %	2,4 %	9,2 %	2,2 %	7,0 %

Fuente: UNESCO / OREALC (2008)

En el marco del Serce, se publicaron aportes para la enseñanza de la lectura, matemática y ciencias, sobre la base de los resultados de las evaluaciones¹³. En la comparación entre países, el Perú se encontró por debajo de la media en todas las evaluaciones, salvo matemática en sexto grado, donde nuestro promedio fue similar a la media regional. El Perú fue el país donde hubo mayor distancia en el rendimiento entre sus estudiantes urbanos y rurales. Este es un dato que en alguna medida muestra la inequidad de nuestro sistema educativo, pues, como se sabe, las zonas rurales suelen concentrar población relativamente pobre e indígena, pero la comparación entre países es débil en lo metodológico, pues cada uno definió urbano y rural usando criterios locales.

Además de las evaluaciones regionales del Llece, el Perú ha participado en dos rondas de la evaluación PISA. De los cuatro ciclos con reportes publicados (2000, 2003, 2006 y 2009), el Perú participó el 2000¹⁴ y el 2009, y nuevamente ha participado el 2012 (aunque los resultados no están disponibles al momento de escribir el presente reporte). En el 2000 y el 2009, el énfasis se puso en la evaluación de comprensión lectora (como se mencionó antes, también se evaluaron habilidades matemáticas y en ciencias). Los resultados son desalentadores: la posición relativa del Perú ha sido la última (2000) o entre los últimos (2009). Sin embargo, se debe considerar que la mayor parte de participantes en esta evaluación son países miembros de la OCDE, es decir, con mayores niveles de riqueza que el Perú.

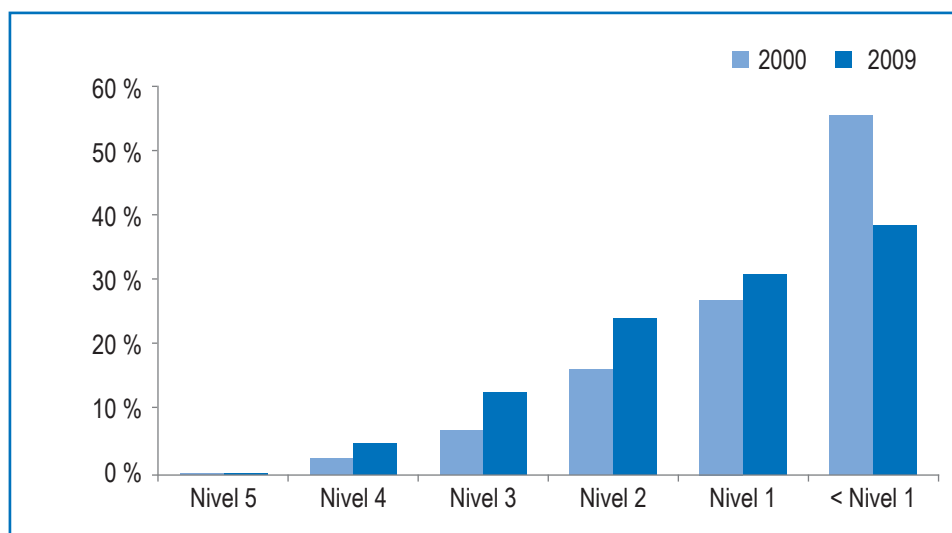
12 Los informes, cuestionarios y bases de datos del 2001 están disponibles en <http://umc.minedu.gob.pe/?p=207> y los del 2004 en <http://umc.minedu.gob.pe/?p=211>

13 En <http://www.llece.org/public/content/view/69/13/lang/es/> se consigue abundante información sobre el Laboratorio, sus evaluaciones, instrumentos, bases de datos y publicaciones.

14 Si bien la mayoría de países participó en PISA el año 2000, el Perú y algunos otros administraron los instrumentos el 2001.

Más importante que la posición relativa del país es analizar lo que los datos sugieren en cuanto a las habilidades de los estudiantes. El gráfico 1 presenta una comparación del porcentaje de estudiantes por nivel de rendimiento en las dos evaluaciones:

Gráfico 1: Resultados de las evaluaciones PISA en comprensión lectora



Fuente: OECD (2010)

Como se puede apreciar, el porcentaje de los estudiantes en el nivel más bajo se redujo considerablemente el 2009. **De hecho, los estudiantes peruanos mejoraron en promedio más que cualquier otro país en comprensión de lectura en PISA entre el 2000 y el 2009, lo cual no quita que la mayoría de los estudiantes peruanos sigue sin llegar a niveles aceptables de rendimiento.** En términos de equidad, para el caso del Perú, se encontró que las variables socioeconómicas de los estudiantes explicaban el rendimiento en mayor proporción que en cualquiera de los otros países participantes. Este es un indicador más de la alta inequidad de la educación peruana. Los resultados del Perú el 2009 en ciencia y matemática fueron similares a los de lectura, es decir, mostraron notables carencias de la gran mayoría de nuestros estudiantes. Al igual que Llece, la OCDE ha producido una gran cantidad de reportes de factores asociados al rendimiento y sugerencias para política¹⁵. El CNE confía en que, en esta oportunidad, el uso que se dé a la información de PISA 2012, esperada para diciembre del 2013, vaya más allá de la posición relativa de los estudiantes peruanos y que se centre en las lecciones que se podrían sacar en lo pedagógico y en el desarrollo de políticas.

Evaluaciones censales

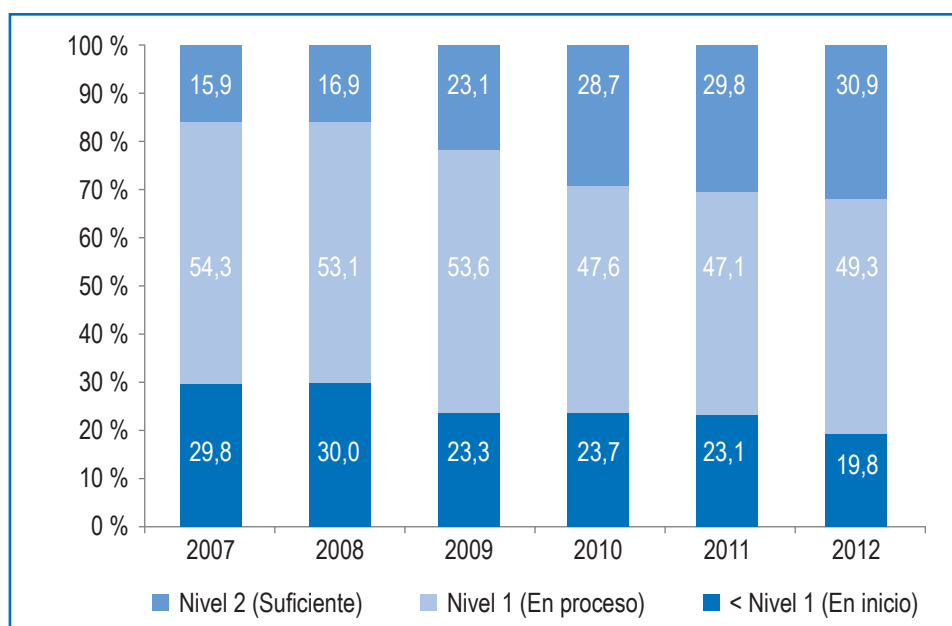
En esta sección, se realizará un análisis detallado de los resultados de las evaluaciones censales, dada su reciente notoriedad como indicador de la calidad y equidad del sistema educativo. Como se dijo antes, según el Marco de Trabajo de la ECE (2009), se excluye de la evaluación censal las instituciones educativas con menos de cinco estudiantes en segundo grado de primaria; por otro lado, las identificadas como Instituciones Educativas de Educación Intercultural Bilingüe (EIB) no se evalúan en segundo, sino en cuarto grado. En cuanto a las instituciones educativas EIB, entendemos que la identificación de las mismas es difícil de realizar usando las bases estadísticas nacionales. Esto ha generado problemas para la evaluación y también para asegurar que las muestras

¹⁵ Ver <http://www.oecd.org/pisa/pisaenespaol/>.

de cuarto grado en diferentes años sean comparables. La evaluación de escuelas EIB en cuarto y no en segundo grado se debe a que, recién en el grado superior, estos estudiantes deberían haber empezado a consolidar sus aprendizajes de lectura en lengua materna y castellano. Sin embargo, hay estudiantes con lengua materna indígena en las evaluaciones de segundo grado. Entendemos que el Ministerio de Educación se encuentra actualmente definiendo qué instituciones educativas deben ser consideradas EIB para atender las necesidades y potencialidades educativas de los estudiantes. En general, la EIB es un campo donde todavía hay mucho por avanzar, desde la definición e identificación de las instituciones educativas y estudiantes en bases de datos válidas, y la atención en lengua materna incluso a individuos que estudien en un ambiente predominantemente castellano.

En cuanto a los resultados de segundo grado de primaria, si bien la evaluación llega a casi la totalidad de estudiantes, para garantizar la confiabilidad y representatividad de los resultados se diseñó la Muestra de Control¹⁶ de la Evaluación Censal para Estudiantes, la cual es representativa en el ámbito nacional y para diferentes estratos (por ejemplo, estatal - no estatal). En las instituciones educativas de esta muestra, se establecen algunos mecanismos de control de calidad, como, por ejemplo, asignar los mejores encuestadores, asignar un encuestador adicional para cada institución educativa y hacer un control de calidad adicional al traslado de las respuestas de los estudiantes. Los resultados que se muestran a continuación corresponden a la Muestra de Control¹⁷. En cuanto a los niveles de rendimiento, el nivel 2 es considerado “satisfactorio”, pues los estudiantes logran los aprendizajes previstos por el currículo. El nivel 1 es definido como “en proceso”, y el nivel debajo del 1 es denominado “en inicio”. Como se dijo antes, la UMC utilizó una metodología basada principalmente en el juicio de expertos para definir estos niveles de rendimiento, y se esperaría que todos los estudiantes se ubiquen en el nivel 2. Luego de definir los niveles en la UMC, se tomaron provisiones para poder comparar los puntajes entre años; en términos estrictos, los puntos de corte no son comparables entre lectura y matemática.

Gráfico 2: Nivel de desempeño en comprensión lectora en las evaluaciones censales 2007-2012



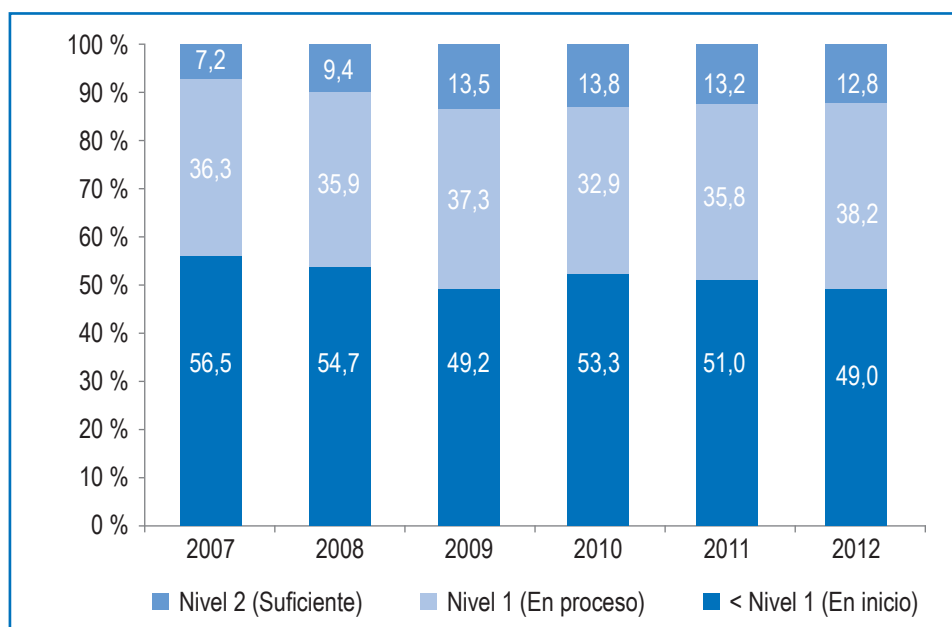
Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2007-2012.

¹⁶ Para mayor información sobre los objetivos y diseño de la evaluación 2012, incluyendo características de la muestra de control ver http://www2.minedu.gob.pe/umc/ece/Marco_de_Trabajo_ECE.pdf

¹⁷ El 2012 la muestra de control tuvo alrededor de 5900 instituciones educativas.

En cuanto a comprensión de lectura (Gráfico 2), ha habido una clara disminución de estudiantes en el nivel por debajo de 1 y un incremento del porcentaje que alcanza el nivel 2. Este último, sin embargo, muestra un avance reducido desde 2010; la disminución del nivel por debajo de 1 ha sido un poco mayor entre el 2011 y 2012. En cuanto a matemática (Gráfico 3), el progreso ha sido bastante menor que en el caso de comprensión de lectura, aunque, si se compara el año 2012 con el año 2007, ha habido un incremento en el porcentaje de estudiantes que alcanza el nivel 2 y una disminución del porcentaje de estudiantes con logros por debajo del nivel 1. Es claro, sin embargo, que la gran mayoría de los estudiantes no logra resolver preguntas de matemática del nivel que se esperaría dado el currículo. El CNE resalta que el Estado se haya comprometido a seguir publicando resultados comparables en el tiempo, a pesar de que a menudo no han sido favorables para el gobierno de turno.

Gráfico 3: Nivel de desempeño en matemática en las evaluaciones censales 2007-2012

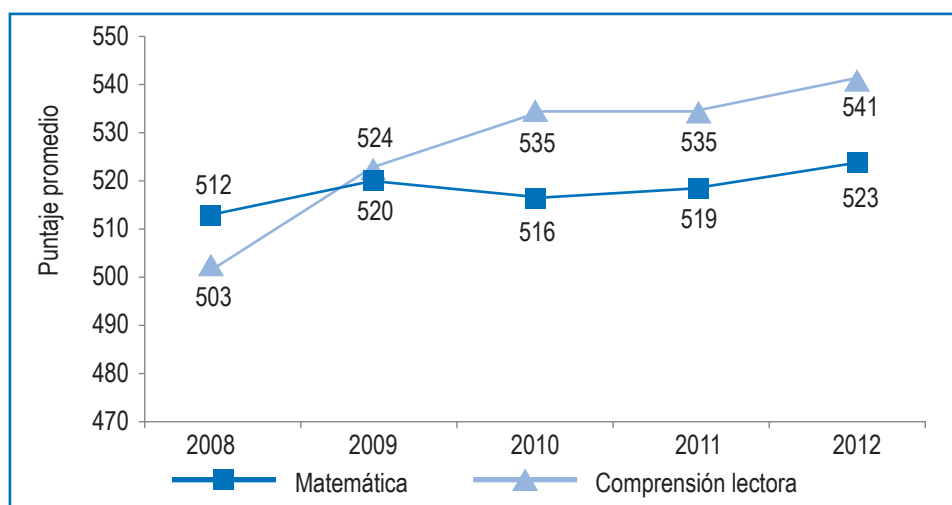


Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2007-2012.

Los anteriores resultados, al reportar tres niveles de rendimiento, ilustran con claridad los niveles de rendimiento a través del tiempo, sin embargo, no permiten aclarar si en promedio se ha mejorado. Para responder esta interrogante, a continuación presentamos algunos resultados del período entre 2008 y 2012 obtenidos de las bases de datos que provee la misma UMC¹⁸. Para los análisis que siguen, en vez de usar las categorías de rendimiento que reporta la UMC, usamos los puntajes de los estudiantes (que son los que permiten generar las categorías) y, sobre esa base, calculamos los promedios. Para estos análisis, la UMC convirtió los puntajes de la primera evaluación para contar con promedio de 500 y una desviación estándar de 100. Este es un procedimiento usado comúnmente en pruebas internacionales, que permite analizar la evolución de los resultados en términos de desviaciones estándar.

18 Ver <http://umc.minedu.gob.pe/?cat=10>

Gráfico 4: Puntaje promedio en las evaluaciones censales 2008-2012

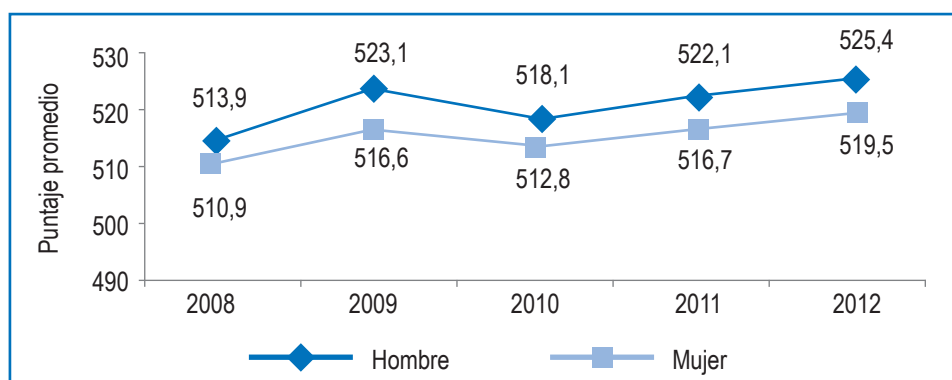


Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2007-2012

Los resultados muestran que en lectura hubo un incremento notable entre el 2008 y el 2009, seguido luego de un incremento menor (entre 2010 y 2011 no hubo crecimiento); en total, desde el 2008 hasta el 2012 ha habido un incremento de 38 puntos. En matemática ha habido un incremento bastante menor entre el 2008 y el 2012 (11 puntos), y no ha habido ningún progreso entre el 2009 y el 2011. **Sugerimos que, en el futuro, la UMC presente los resultados nacionales de las dos formas, en categorías y como promedios, como se hace con todas las evaluaciones internacionales.** Esto permitirá mejores análisis de los resultados y sus implicancias.

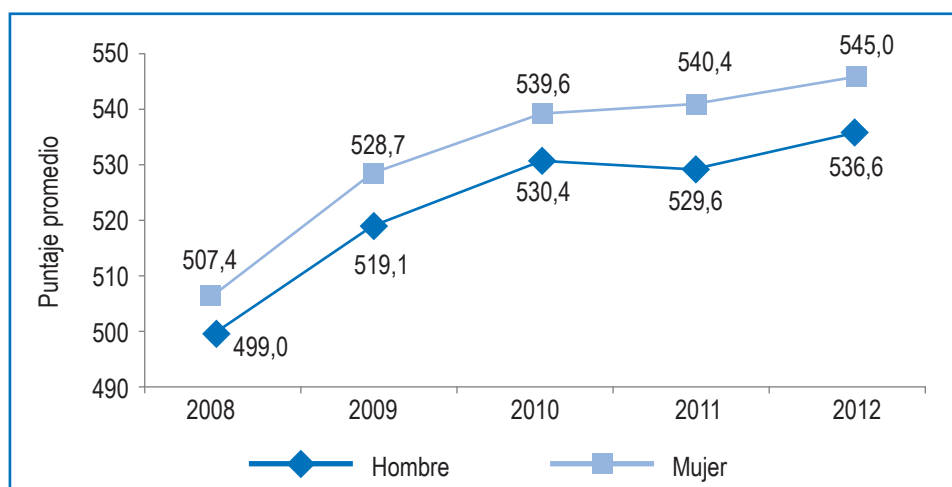
Un tema crucial para la educación peruana es la inequidad, que en este caso se refleja en las diferencias en rendimiento entre grupos de estudiantes. La UMC ha venido presentado estas diferencias, o brechas, de acuerdo con la clasificación en tres niveles. Si bien esta forma de presentación es correcta técnicamente, no provee el promedio para cada grupo. A continuación, presentamos los promedios por grupo (es decir, usamos la misma base de datos que se usó para el gráfico 4). En primer lugar, se presentan los resultados para hombres y mujeres. El gráfico 5 muestra que en matemática hay una ligera diferencia a favor de los hombres, que se ha mantenido más o menos constante durante el tiempo. En lectura, en cambio, el rendimiento de las mujeres ha sido de manera constante superior al de los hombres. Este es un resultado común en evaluaciones internacionales. En todo caso, las brechas entre hombres y mujeres son menores que las que se observan en las siguientes dimensiones:

Gráfico 5. Puntaje promedio en matemática según género



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

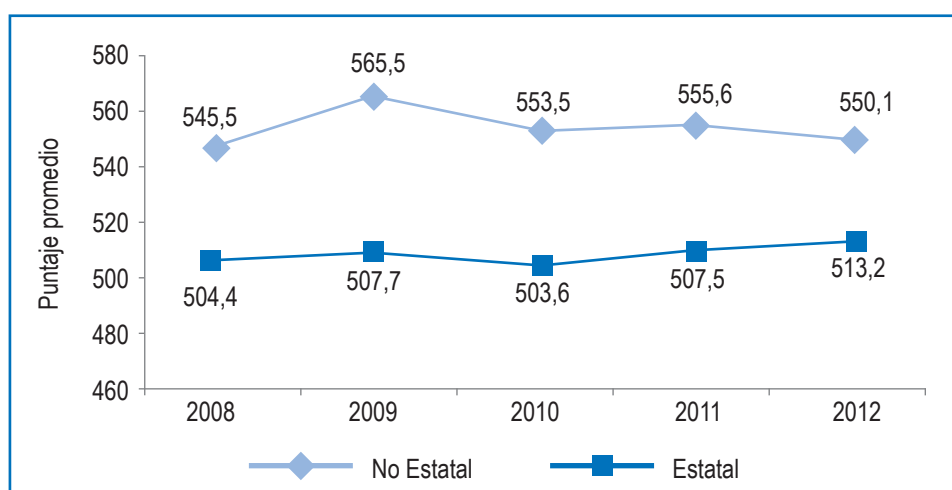
Gráfico 6: Puntaje promedio en comprensión de lectura según género



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

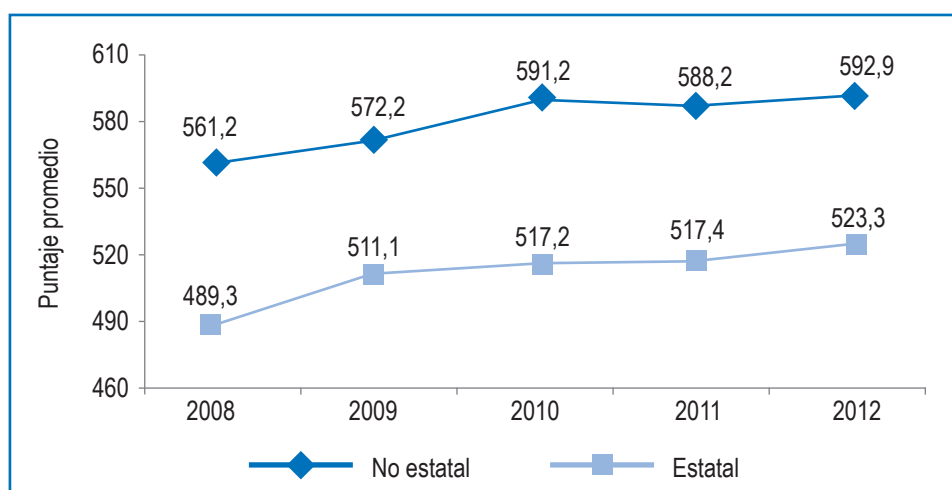
Los gráficos 7 y 8 presentan las brechas por gestión de la institución educativa. Tanto en matemática (Gráfico 7) como en comprensión (Gráfico 8) se nota una brecha a favor de los no estatales. En matemática, esta brecha era de, aproximadamente, 41 puntos el 2008 y se ha reducido a 37 puntos el 2012. En lectura, la brecha era aún mayor el 2008, alrededor de 72 puntos, y prácticamente no se ha reducido hacia el 2012.

Gráfico 7: Puntaje promedio en matemática según gestión de la escuela



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Gráfico 8: Puntaje promedio en comprensión de lectura según gestión de la escuela



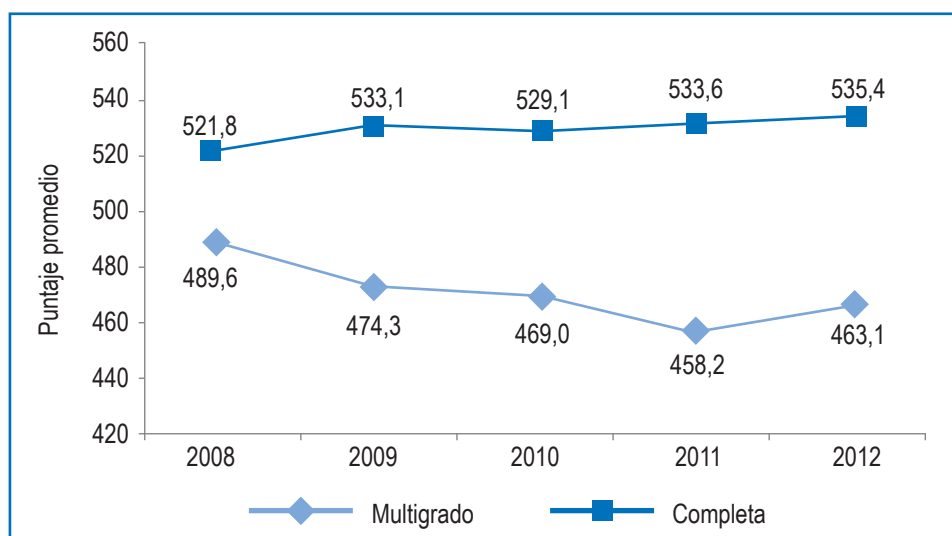
Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Una interpretación rápida, y claramente errada, sería que la educación no estatal es de superior calidad que la estatal. La investigación internacional y la local muestran que uno de los principales factores explicativos del rendimiento son las características socioeconómicas del estudiante y su familia. Así, al ser la educación no estatal de paga, concentra a las familias con mayores recursos. Identificar cuánto de la diferencia entre el logro de aprendizaje de los estudiantes que asisten a instituciones educativas estatales y los que asisten a no estatales se debe a la calidad de la gestión, y cuánto se debe a las características del estudiante y su familia, o a la interacción entre estas, es prácticamente imposible. Por lo demás, obviamente hay mucha variabilidad entre instituciones educativas tanto al interior del grupo estatal como del no estatal, lo cual sugiere que la calidad pedagógica y características de los estudiantes también son muy diversas al interior de ambos grupos.

Finalmente, presentamos el promedio de los estudiantes que asisten a escuelas polidocentes completas y el de aquellos que asisten a escuelas multigrado. Pensamos que esta distinción es relevante en tanto se aproxima a diferencias por ubicación geográfica (las escuelas multigrado suelen ser rurales; las completas, urbanas) y de trabajo pedagógico (a menudo para los docentes que trabajan en escuelas multigrado es un reto educar simultáneamente a estudiantes de varias edades y grados sin contar con materiales o entrenamiento específico para este tipo de trabajo)¹⁹. En matemática (Gráfico 9), la brecha el 2008 era de alrededor de 32 puntos; en 2012, esta brecha se ha ampliado a alrededor de 72 puntos. En comprensión de lectura, la brecha el 2008 era de alrededor de 56 puntos, y, el 2012, se había ampliado a 84 puntos. En ambos casos, sin embargo, hubo una ligera disminución de la brecha entre el 2011 y el 2012.

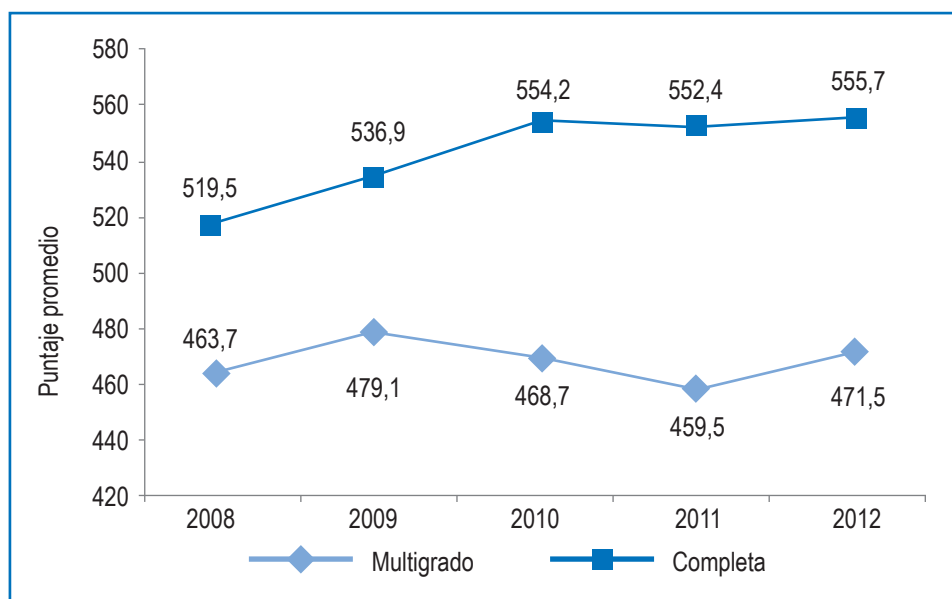
¹⁹ En el presente informe no comparamos instituciones educativas urbanas y rurales, ya que el 2010 la Unidad de Estadística Educativa del Ministerio de Educación reclasificó como urbano a un conjunto importante de centros poblados ubicados en la periferia de las ciudades y que previamente habían sido considerados como rurales. La UMC debía usar este mismo padrón, pero al hacerlo la comparación de las evaluaciones con años previos no es válida.

Gráfico 9: Puntaje promedio en matemática según característica de la escuela



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Gráfico 10: Puntaje promedio en comprensión de lectura según característica de la escuela



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Los anteriores datos refuerzan la importancia de la inequidad en los aprendizajes como factor crucial para el desarrollo de la educación peruana. Sabemos que la actual gestión ha puesto este tema como prioritario en sus políticas y entendemos que estamos enfrentando desigualdades históricas que seguramente será muy difícil revertir. Confiamos, sin embargo, en ver los primeros resultados pronto y **sugerimos que, en el futuro, las brechas se presenten, tal como se ha hecho hasta ahora, como porcentajes de estudiantes que logran niveles de rendimiento, pero también como puntajes promedio en, al menos, las dimensiones reportadas antes en este documento.** Quisiéramos notar, sin embargo, que, salvo por sexo, todas las demás categorías presentadas se basan en grupos de escuelas. Como se ha mencionado en otras partes, las características individuales de los estudiantes se encuentran asociadas al rendimiento. Así, **sugerimos que, en futuros**

reportes, se presenten promedios de grupos de estudiantes en, al menos, dos categorías adicionales al sexo: alguna medición del nivel socioeconómico (por ejemplo, a través de la educación de la madre o de ambos padres) y otra de la lengua materna.

Una de las ventajas de la evaluación en segundo grado es que permite estimar promedios regionales, aunque no todos los años cada región haya cumplido con los estándares necesarios para que la muestra fuera representativa (en cuyo caso la UMC no reportó resultados para la región en cuestión, ya que el error muestral sería inaceptablemente alto). El cuadro 4 muestra los promedios de las regiones en comprensión de lectura y matemática entre el 2008 y el 2012. El 2012 se logró tener promedios por primera vez para todas las regiones.

Cuadro 4: Puntaje promedio en las evaluaciones censales por región 2008-2012

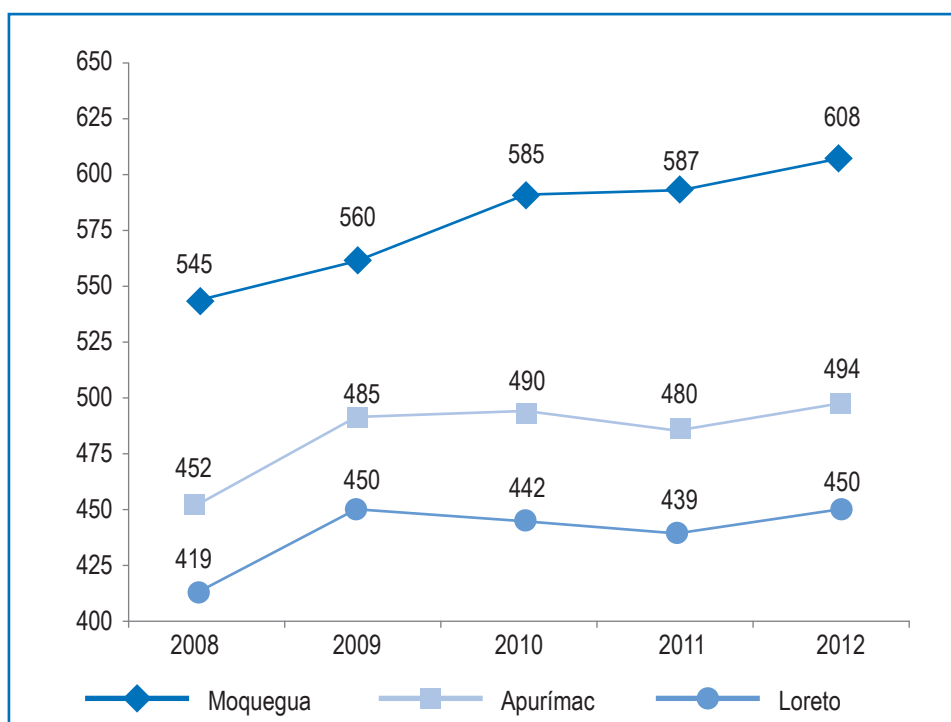
Región	Matemática					Comprensión lectora				
	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012
Amazonas	515	509	499	515	519	488	509	507	514	517
Áncash	501	509	505	501	493	485	512	510	510	513
Apurímac	479	483	480	471	488	452	485	490	480	494
Arequipa	543	567	566	560	560	551	563	586	585	589
Ayacucho					480					497
Cajamarca					502					506
Callao	532	536	540	541	555	537	549	564	570	579
Cusco	503	502	506	502	501	490	509	520	513	515
Huancavelica			519	490	504			505	488	500
Huánuco	490	473	486	490	479	463	481	491	493	490
Ica	534	566	568	551	550	527	547	568	564	561
Junín	522	541	519	528	528	508	532	539	537	541
La Libertad	515	539	516	521	528	504	532	530	531	543
Lambayeque	518	527	534	535	525	512	533	553	551	549
Lima Metropolitana	528	548	545	553	560	542	557	578	578	587
Lima provincias	523	535	514	526	536	512	535	542	549	553
Loreto	433	427	425	421	425	419	450	442	439	450
Madre de Dios	504	487	489	507	501	496	505	513	520	526
Moquegua	544	577	573	587	619	545	560	585	587	608
Pasco	526	522	514	500	514	498	525	529	515	524
Piura	509	511	511	520	526	498	520	533	534	539
Puno		524		498	503		521		509	517
San Martín	478	472	476	486	491	470	488	502	499	507
Tacna	552	567	592	582	614	553	559	592	583	601
Tumbes	508	522	534	523	525	504	523	547	535	539
Ucayali	459	447	455	459	466	456	474	490	491	499

Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012. Los espacios en blanco indican regiones que ese año no tuvieron una muestra suficiente que permitiera reportar resultados representativos.

Del cuadro anterior se desprende que hay una gran variabilidad en la evolución de los promedios regionales. En matemática, las regiones que han subido más que el promedio desde el 2008 son Arequipa, Callao, Ica, La Libertad, Lima Metropolitana, Lima provincias, Moquegua, Piura, San Martín, Tacna y Tumbes. En comprensión de lectura, las regiones que más han subido son Amazonas, Apurímac, Callao, La Libertad, Lima provincias, Moquegua, Piura, Tacna y Ucayali. Con algunas excepciones, el panorama es de mejora del rendimiento principalmente en regiones de la Costa, que tienen niveles relativamente bajos de pobreza y donde predomina el castellano. Aun así, el análisis de casos de alto rendimiento y alto incremento, por ejemplo, Moquegua y Tacna, aunque también el Callao, La Libertad, Lima provincias y Piura, arrojaría seguramente lecciones interesantes de política a nivel regional y de las instituciones educativas.

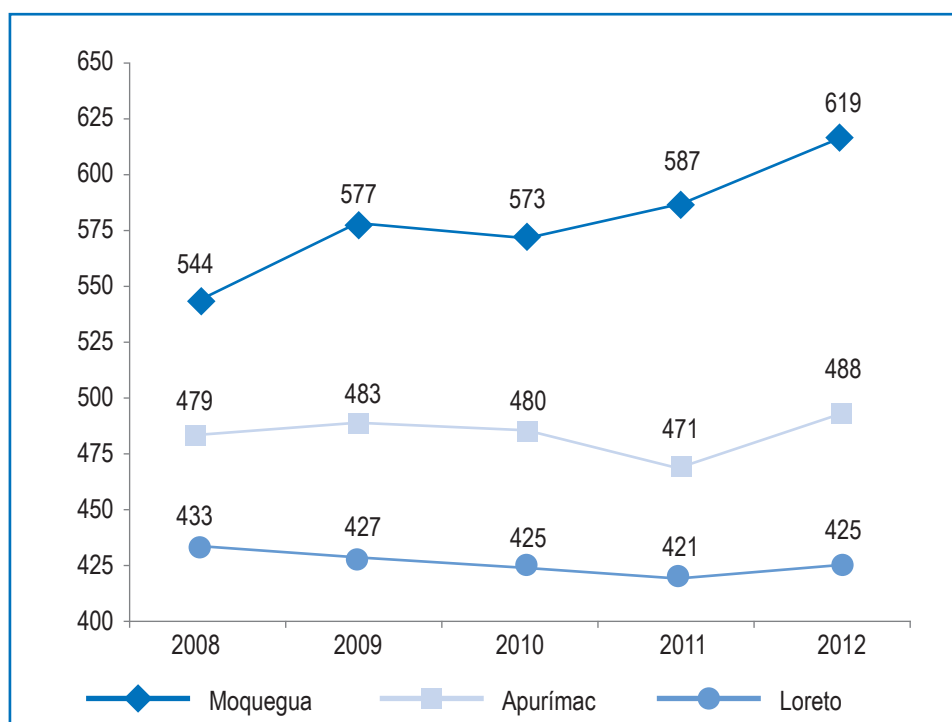
El Ministerio de Educación reconoció recientemente los mayores incrementos en rendimiento, dados en Apurímac, Moquegua y Tacna entre el 2011 y el 2012. Para entender la evolución histórica de las dos primeras regiones, en el siguiente gráfico se incluyen sus resultados (los resultados de Tacna son similares a los de Moquegua). Como se puede ver, el progreso de Moquegua es casi constante, mientras que el de Apurímac ha sido más irregular. El CNE confía en que el rendimiento de regiones con altos niveles de pobreza, como Apurímac, continúe incrementándose en los próximos años, sin embargo, para futuros reconocimientos sugerimos que se consideren las tendencias en el incremento en varios años y no solo el incremento entre un año y otro. Presentamos esta sugerencia porque consideramos que, para determinar si una gestión educativa prioriza aprendizajes, es mejor considerar un incremento sostenido a lo largo de los años en lugar de un incremento volátil de un año al siguiente. En los gráficos 11 y 12 se incluye también el promedio de Loreto, que si bien ha mejorado en lectura, sigue siendo el más bajo del país en ambas áreas. Esta región probablemente necesite una acción intersectorial urgente para mejorar los niveles de aprendizaje de los estudiantes, tarea en la que sabemos está comprometida la actual gestión.

Gráfico 11: Puntaje promedio de regiones seleccionadas en comprensión lectora



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Gráfico 12: Puntaje promedio de regiones seleccionadas en matemática

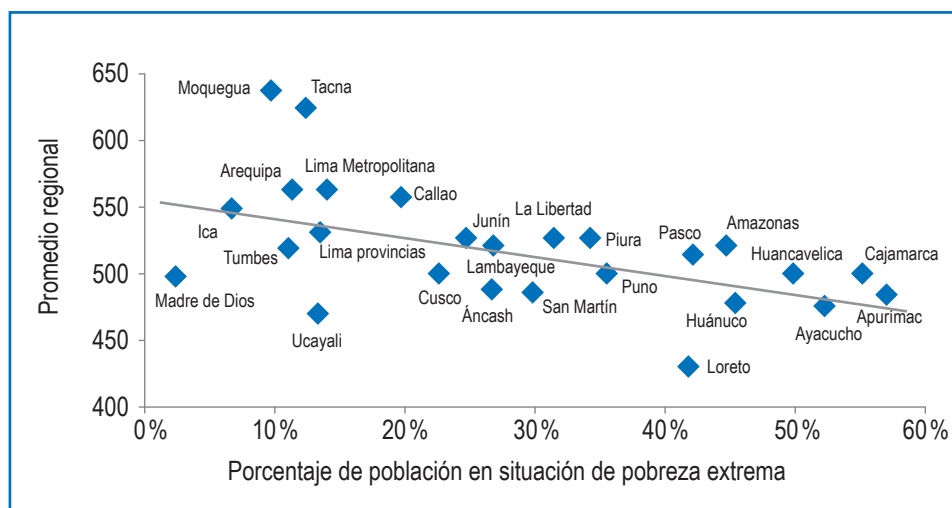


Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012.

Como se ha dicho en otras partes del presente documento, estudios previos muestran una asociación clara entre rendimiento y pobreza. Los gráficos 13 y 14 presentan evidencia al respecto. Así, se presenta el porcentaje de pobreza a nivel de la región²⁰, y el promedio de la región en la evaluación censal 2012 en matemática (Gráfico 13) y lectura (Gráfico 14). En los gráficos siguientes, cada uno de los rombos representa una región y expresa dos valores: por su posición en el eje horizontal, indica el porcentaje de población en situación de pobreza; por su posición en el eje vertical, el puntaje logrado en la Evaluación Censal de Estudiantes, tanto en matemática (Gráfico 13) como en comprensión lectora (Gráfico 14). La relación es claramente negativa: a mayor porcentaje de pobres en la región, menor rendimiento.

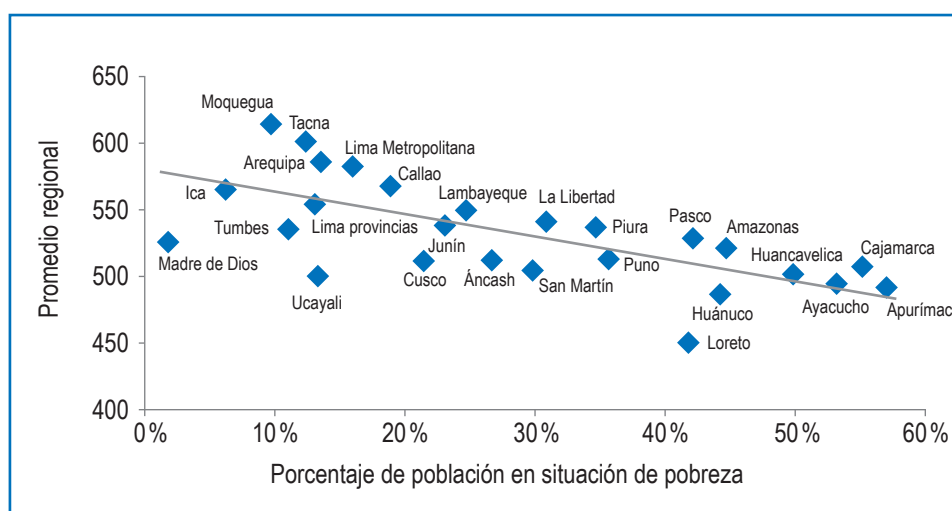
²⁰ El dato de pobreza se obtuvo del módulo Sumarias de la Encuesta Nacional de Hogares 2012. Las familias con un ingreso menor a la línea de pobreza establecida son consideradas pobres.

Gráfico 13: Relación entre el promedio regional en matemática y el porcentaje de pobreza (2012) ($r = -0,54$)



Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012 y ENAHO 2012

Gráfico 14: Relación entre el promedio regional en comprensión y el porcentaje de pobreza (2012) ($r = -0,67$)

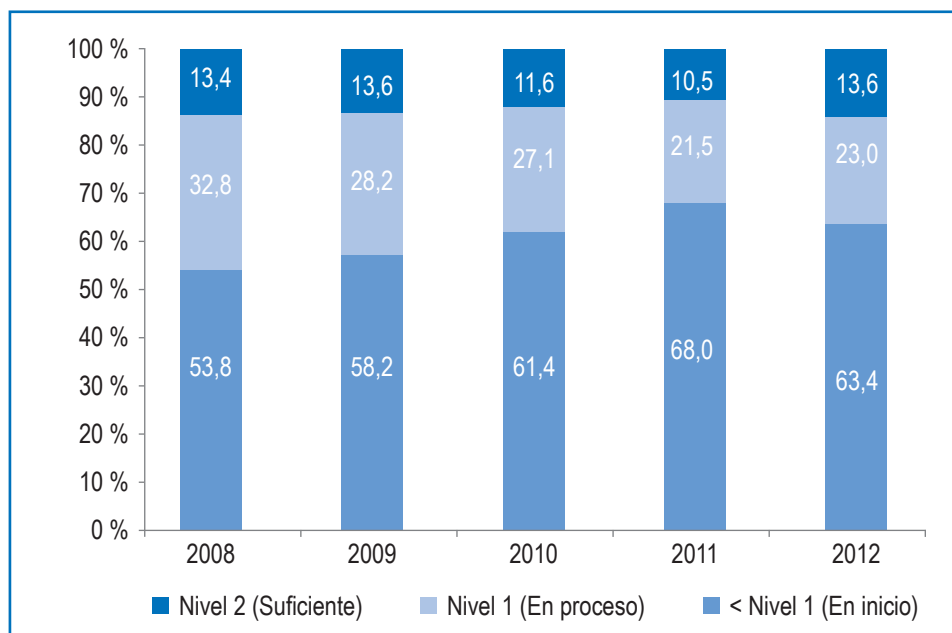


Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2008-2012 y ENAHO 2012.

Después de examinar los gráficos anteriores, con un criterio de equidad se esperaría que el gasto público en educación fuera mayor por estudiante en las regiones con menor rendimiento. Los siguientes gráficos presentan la relación entre el promedio regional en la evaluación censal 2012 y el gasto público por estudiante²¹ en educación primaria en el 2011 (último año disponible). La relación en ambos casos es positiva: a mayor rendimiento, mayor gasto. Pensamos a partir de estos datos y datos similares de otros estudios que **habría que diseñar un nuevo sistema de asignación del presupuesto, que incremente los montos disponibles por estudiante para las regiones, provincias y distritos con menor rendimiento y mayores niveles de pobreza.**

21 El gasto público por estudiante fue obtenido de la base de Estadística de Calidad Educativa como resultado de dividir el gasto público en primaria, luego de excluir las transferencias a hogares no gastadas en instituciones educativas, entre el número de alumnos matriculados en instituciones educativas públicas en educación primaria.

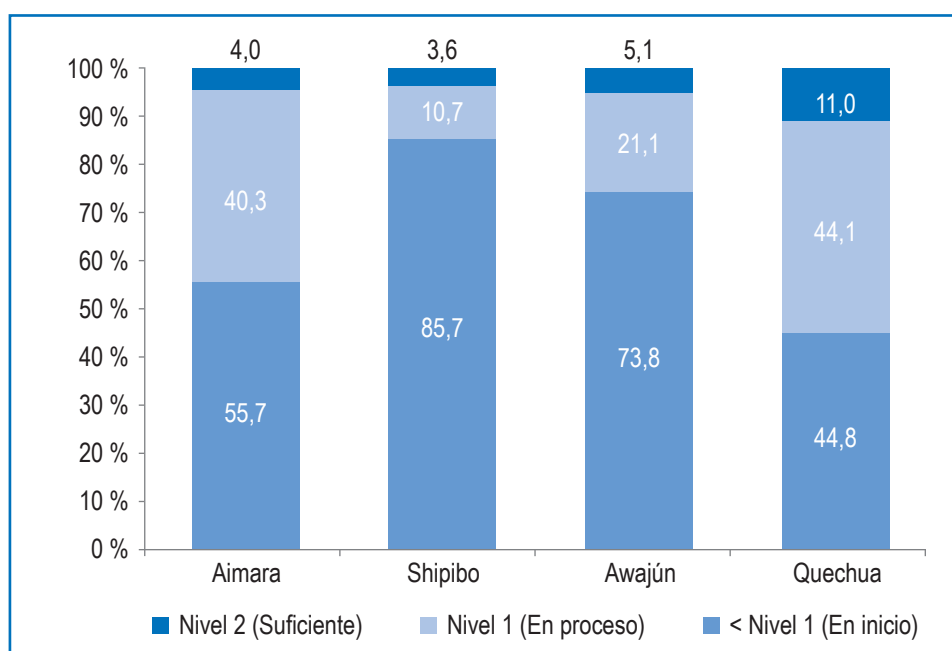
Gráfico 17: Comprensión lectora en castellano como segunda lengua



Fuente: Resultados Evaluaciones Censales Estudiantiles 2008-2012.

El gráfico 18 presenta los resultados de los estudiantes indígenas en comprensión de lectura en su lengua materna en la evaluación del 2012. Como se recordará, todos estos estudiantes se encontraban en escuelas con el modelo de educación intercultural bilingüe, por lo que deberían aprender a leer y escribir en lengua materna, sin embargo, son muy pocos los estudiantes que logran un nivel suficiente en la prueba. De entre los cuatro grupos, los resultados más altos los obtienen los estudiantes cuya lengua materna es el quechua.

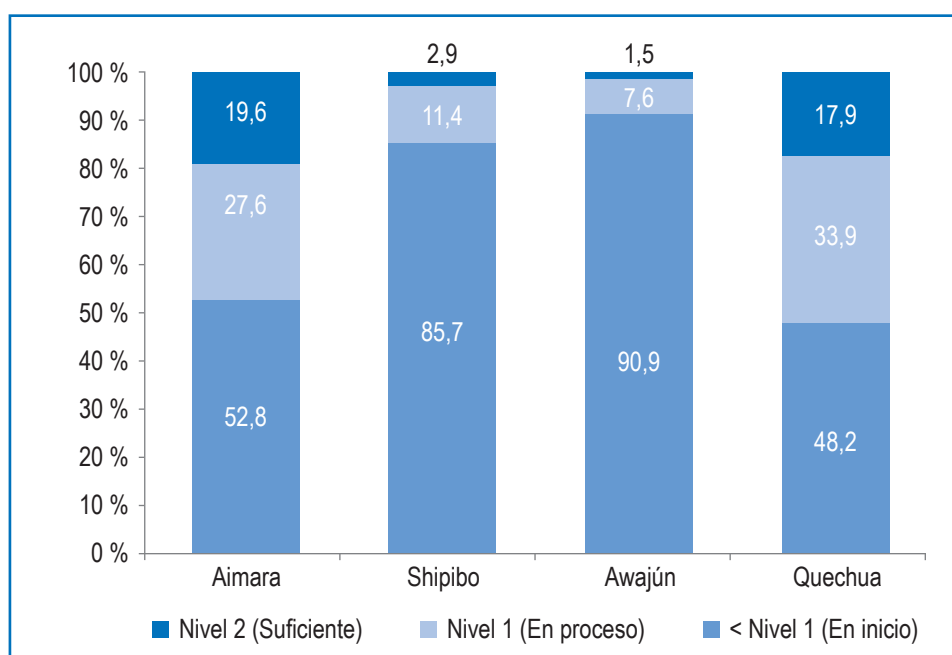
Gráfico 18: Comprensión lectora en lengua originaria (2012)



Fuente: Resultados Evaluaciones Censales Estudiantiles 2012.

Finalmente, a estos mismos estudiantes se les evaluó en comprensión de lectura en castellano como segunda lengua. Los estudiantes que lograron el nivel satisfactorio son una minoría en todos los casos; además, hay notables diferencias entre grupos étnicos. Los estudiantes aimara y quechua tienen rendimientos muy por encima de los estudiantes shipibo y awajún; la situación es dramática para estos dos grupos, con al menos 85% de los estudiantes en el nivel más bajo de lectura. **El CNE expresa su honda preocupación por el bajo rendimiento de los estudiantes indígenas en pruebas de comprensión lectora en lengua materna y castellano como segunda lengua, y sugiere continuar estas evaluaciones y hacer de estos estudiantes una prioridad en el presupuesto y el diseño de programas.**

Gráfico 19: Comprensión lectora en castellano como segunda lengua, según lengua originaria (2012)



Fuente: Resultados Evaluaciones Censales Estudiantiles 2012.

En cuanto al uso de la información, la UMC elabora diversos reportes con base en las ECE. Estos están dirigidos a la institución educativa, a los docentes (con datos, ejemplos y sugerencias para lectura y matemática) y a los padres de familia, con los resultados de cada estudiante evaluado. Adicionalmente, hay informes para regiones y para las UGEL. Específicamente, para cuarto de primaria, hay además reportes para el director y el docente que explican las evaluaciones, los resultados y que sugieren alternativas pedagógicas en EIB. Los informes no solo se distribuyen en impreso, sino que se cuenta con el *Sistema de Consulta de Resultados de la Evaluación Censal de Estudiantes*²². Se trata, sin duda, de un recurso de gran potencial para proporcionar información específica a los principales actores educativos. Sin embargo, nada de lo anterior es relevante si los usuarios no reciben y usan la información. Al respecto, la Encuesta Nacional a Instituciones Educativas de Nivel Inicial y Primaria (Enedu) del 2012 muestra que casi 98% de las instituciones educativas reportan haber recibido los informes²³. Si bien 97% de los directores encuestados piensan que el informe es útil para la toma de decisiones, solo el 62% de los encuestados manifiesta que comprende los resultados de la Evaluación Censal. Luego, solo el 73% de los directores reporta haber entregado los informes de la evaluación censal a sus docentes. Finalmente, casi el 82% de los directores reportó haber entre-

²² Ver http://sistemas02.minedu.gob.pe/consulta_ece/publico/index.php.

²³ Los datos de la ENEDU que se presentan fueron facilitados por personal del Ministerio de Economía y Finanzas, a quienes estamos agradecidos.

gado los informes a los padres de familia. Lo preocupante en este caso es que para zonas rurales es menos probable que hayan entregado el reporte a padres que en zonas urbanas (78 % versus 88 % respectivamente). Con los reportes, como con cualquier instrumento educativo, se debe procurar evitar se fomente mayor inequidad. Esta tarea no debe ser necesariamente asignada a la UMC, sino pensada en un conjunto de iniciativas para promover mayor equidad educativa.

Finalmente, en cuanto a utilización de información, es muy poco lo que ha publicado el MED respecto a los factores asociados al rendimiento en las evaluaciones censales. Los motivos posiblemente sean dos: los cuestionarios de factores asociados son muy breves (esto es común en evaluaciones censales, dado el gran requerimiento logístico para poder contar con pruebas administradas de manera estandarizada), y el escaso tiempo disponible del equipo en la UMC para realizar análisis adicionales a los que se necesitan para producir pruebas y elaborar la diversidad de informes referida antes. El último documento de trabajo que hemos encontrado publicado en el portal web de la UMC es del 2006, que es justamente cuando empezaron las evaluaciones censales. Sin embargo, la UMC ha estado en general abierta a proporcionar los datos disponibles a investigadores de diferentes instituciones, los que han producido en algunos casos informes notables con los datos. Se trata de una tarea, sin embargo, que requeriría ser profundizada pues puede ser muy sugerente para el diseño o revisión de políticas. Volveremos sobre este tema más adelante.

BALANCE Y RECOMENDACIONES

En el CNE reconocemos que es mucho lo que se ha avanzado en cuanto a la evaluación estandarizada del rendimiento de los estudiantes. La instancia responsable dentro del Ministerio de Educación, la Unidad de Medición de la Calidad Educativa (UMC), ha adquirido notable experiencia y capacidad técnica, con lo cual sus resultados son en general percibidos como confiables y válidos. Esto es positivo en una sociedad que promueve la rendición de cuentas, y que tiene gran interés en la educación como instrumento para el desarrollo de sus ciudadanos y el país. Sin embargo, consideramos que es todavía mucho lo que se podría avanzar en cuanto a las prácticas evaluativas y su utilización. A lo largo del presente documento, hemos dado algunas opiniones y sugerencias sobre el sentido y usos de la evaluación; por ejemplo, **nuestra preferencia por mantener un enfoque de pruebas con base en criterios y de bajas consecuencias**. A continuación presentamos sugerencias adicionales sobre las evaluaciones estandarizadas:

1. Sobre las expectativas de aprendizaje. Hasta hace poco, en los sucesivos currículos nacionales peruanos, faltaba precisión sobre lo que los estudiantes deberían saber y poder hacer en las diferentes áreas a lo largo de su escolaridad (es decir, faltó definir lo que en la literatura se denomina “estándares de desempeño”). Las evaluaciones de la UMC de alguna forma han señalado, en la práctica, cuáles eran esos estándares. El Ipeba ha desarrollado recientemente los Mapas de Progreso, que se orientan a estos objetivos, y aún más recientemente el Ministerio de Educación ha buscado completar el sistema curricular con dos documentos complementarios a los Mapas: el Marco Curricular, que explicita los aprendizajes fundamentales, y las Rutas de Aprendizaje, que sugieren pautas metodológicas para la enseñanza de los aprendizajes fundamentales y los estándares contenidos en las Rutas. **El CNE sugiere que, para que este sistema curricular sea una herramienta eficaz para promover los aprendizajes, es necesario establecer coherencia entre los tres documentos, pero también con las evaluaciones censales, las políticas de textos y la formación y desarrollo profesional docente.** De acuerdo con múltiples planteamientos internacionales, es en la sinergia de todos estos instrumentos en donde reside el potencial de aprendizaje de los estudiantes²⁴. Lo que observamos en la actualidad con preocupación es un desarrollo parcial del sistema curricular y, donde no hay nuevos elementos, los docentes deben seguir con el Diseño Curricular Nacional y con textos que obviamente no están alineados con el nuevo sistema.

No se debe olvidar, sin embargo, que el currículo como documento señala apenas una pretensión oficial de lo que debería ocurrir en las aulas. Diversos autores han realizado análisis de lo que ocurre al interior de los salones y han encontrado que se pierde mucho tiempo y que los autores subutilizan los materiales disponibles y cubren una fracción de lo que prescribe el currículo (ver, por ejemplo, Zambrano, 2002; Cueto, Guerrero, Zapata, y Freire, 2013). **Así, sugerimos crear un sistema de monitoreo de la calidad de los procesos pedagógicos en las aulas que permita retroalimentar el currículo para hacerlo una herramienta eficaz en el incremento de aprendizajes.**

24 Ver por ejemplo (OECD, 2013).

2. Sobre los planes de evaluación. En el CNE celebramos los recientes anuncios del Ministerio de Educación respecto de que se va a ampliar el espectro de evaluación para cubrir ciudadanía y ciencias al final de la primaria y la secundaria, en adición a lectura y matemática, pero pensamos que esto todavía es insuficiente. **El CNE sugiere que el Ministerio de Educación desarrolle y discuta un plan de evaluaciones de largo plazo que prevea cubrir todos los aprendizajes fundamentales previstos en el nuevo Marco Curricular, sin que esto signifique que se evalúen todos los grados.** De manera rotativa se podrían cubrir algunos grados y de manera sucesiva y planificada todos los aprendizajes fundamentales, aunque, con mayor frecuencia, los de las áreas instrumentales para el aprendizaje (v. gr. lectura y matemática).

Como parte del plan, sugerimos planificar evaluaciones longitudinales de un conjunto de estudiantes. Tal evaluación permitiría estimar de mejor manera la contribución del sistema educativo a los aprendizajes de un grupo de estudiantes a lo largo del tiempo. La UMC ha realizado un primer e interesante esfuerzo de evaluación longitudinal, pero solo en Lima y con escuelas públicas. Una muestra más variada a escala nacional permitiría recoger información valiosa para el desarrollo de políticas, sobre la base de una mejor estimación del peso de la escuela en la explicación de los incrementos de aprendizaje.

Finalmente, **el plan debería incluir también la participación en evaluaciones internacionales.** En este momento, hay tres instituciones que organizan evaluaciones a escala internacional en las que el Perú ha participado o podría participar: el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa de la Unesco, que se encuentra organizando su tercer estudio en primaria para países de la región; la OCDE, que organiza PISA para evaluar a estudiantes de 15 años; y la Asociación Internacional para la Evaluación (IEA)²⁵, que organiza periódicamente evaluaciones en ciencias, matemática, lectura y ciudadanía, entre otros temas, en primaria y secundaria. En el CNE consideramos que, una vez que se establezca un plan nacional de evaluaciones, habría que considerar qué evaluaciones internacionales podrían complementar las planificadas. Por ejemplo, con Lleece cubrimos primaria y con la OCDE secundaria; de la IEA lo más novedoso podría ser participar en el tercer estudio sobre educación ciudadana o en alguno de sus estudios especiales (por ejemplo, recientemente culminaron el estudio TEDS-M sobre formación docente en matemática alrededor del mundo). Participar en evaluaciones internacionales tiene múltiples ventajas, como por ejemplo conocer y aplicar estándares internacionales para el contexto peruano, poder aprender de nuestros resultados en una perspectiva comparada y fortalecer las habilidades técnicas del equipo evaluador local. Sin embargo, al igual que con cualquier evaluación, desde el inicio debe haber un plan claro de utilización de los resultados de las evaluaciones internacionales.

3. Sobre evaluaciones censales o muestrales. Si bien el CNE reconoce que las evaluaciones censales han sido un gran estímulo para priorizar los aprendizajes, pensamos que **habría que adoptar un esquema de evaluaciones muestrales** acorde con el argumento presentado en el punto anterior de contar con un plan a largo plazo que cubra todos los aprendizajes fundamentales. Las evaluaciones censales requieren grandes inversiones y trabajo dedicado para hacerse bien, a costa de la profundidad de las pruebas y de la realización de otras pruebas y análisis. Para no perder uno de los usos principales de la evaluación censal, el de entregar a cada institución educativa una copia de los instrumentos y sus resultados individuales, **sugerimos que, luego de terminadas las evaluaciones muestrales, se libere una muestra de los ítems con instrucciones para los docentes en cada institución educativa sobre cómo administrarlos; también consideramos conveniente procesar los resultados de modo que se puedan comparar con los resultados nacionales.** Solo se deberían mantener en reserva los ítems necesarios para poder hacer equiparaciones de rendimiento en el tiempo. Finalmente, **las evaluaciones muestrales deberían ser acompañadas de instrumentos de contexto (por ejemplo, cuestionarios a di-**

25 En inglés *International Evaluation Association*, ver <http://www.iea.nl>.

versos actores y observaciones) que permitan caracterizar el contexto en que estudian los estudiantes y establecer la asociación de este con su rendimiento. Estos datos en conjunto podrían ser luego relevantes para hacer análisis de factores asociados al rendimiento que puedan sugerir políticas concretas.

4. **Sobre los usos de las evaluaciones.** Si bien, como se dijo antes, el CNE reconoce la calidad de los informes preparados por la UMC para diversos actores, pensamos que sería necesario diversificar los informes disponibles. En primer lugar, **sugerimos involucrar más activamente en el uso de los reportes de las evaluaciones, primero, a los autores y editoriales de textos escolares, y, segundo, a los formadores y capacitadores de docentes, con el fin de que sus materiales y procedimientos estén alineados con las características de la evaluación.**

En segundo lugar, consideramos pertinente continuar con las políticas de diseminación de los resultados y herramientas de evaluación entre directores, docentes, tomadores de decisiones a nivel central y regional, y padres y madres de familia, además de la opinión pública a través de los medios de comunicación masiva. Sin embargo, de acuerdo con los datos de Enedu mencionados antes hay todavía una subutilización de la información disponible, pareciera que principalmente en zonas rurales.

Para cumplir con todo lo anterior, **el CNE sugiere que la UMC diseñe un plan de comunicaciones y desarrollo de capacidades de los diferentes tipos de usuarios. El plan debe tener como objetivo, primero, diseminar la información; segundo, explicar los principales mensajes con un criterio de equidad que, por ejemplo, se focalice en padres e instituciones educativas en zonas de pobreza, indígenas y estudiantes con menor rendimiento; tercero, alentar a que los diferentes actores desarrollen habilidades para extraer implicancias de política coherentes con los resultados, de modo que se limite así la posibilidad de que se hagan interpretaciones sin fundamento.** En esta línea, alentamos que en los cursos de formación docente y en los de desarrollo profesional se introduzcan cursos o módulos sobre evaluación educativa. También alentamos que alguna universidad cree una especialización en evaluación educativa, dada la relevancia del tema.

Reconocemos el esfuerzo de la UMC en cuanto a comunicaciones, sobre todo con los magníficos reportes mencionados antes, pero de nuevo señalamos que la utilización coherente e intensa de los resultados es lo que en última instancia justifica el gran esfuerzo realizado para medir los niveles de aprendizaje de los estudiantes.

En tercer lugar, si bien reconocemos que la UMC ha estado en general abierta a proveer de información a la comunidad académica, pensamos que esta labor se podría reforzar. **Sugerimos que el Ministerio convoque a concursos de investigación sobre los temas que considere prioritarios, de modo que se asegure que los reportes de investigación tengan una utilidad para políticas.**

5. **Sobre la adaptación de las evaluaciones a poblaciones minoritarias.** La evaluación de estudiantes indígenas en lengua materna y castellano como segunda lengua es una labor que el Ministerio debe continuar. El Ministerio, de manera rotativa, debe incluir todas las lenguas indígenas que sea posible²⁶, de acuerdo con un criterio de equidad. Saludamos que recientemente el Ministerio de Educación haya confirmado que seguirá evaluando a escala nacional a las poblaciones indígenas, pues dado el actual avance técnico en el país, si no lo hace la UMC, se corre el riesgo de que esta práctica se pierda o que la hagan instancias con menor

²⁶ Entendemos que hay algunas lenguas indígenas con muy pocos hablantes, para las cuales sería sumamente complicado técnicamente desarrollar evaluaciones. Lo mínimo que se debería esperar con estos grupos es seguir evaluando a estos grupos en comprensión de lectura en castellano como segunda lengua.

experiencia y capacidad, y esta es una población estudiantil con muy pobres resultados, en la que se hace urgente observar mejoras pronto. **Por otro lado, sugerimos que la UMC considere la evaluación de estudiantes con discapacidad dentro de su plan de evaluaciones y como parte de sus políticas a favor de la equidad.**

6. Sobre los documentos técnicos de apoyo. Todas las evaluaciones internacionales de prestigio cuentan con documentos técnicos que explican los fundamentos y propósitos del estudio, el marco muestral, los procedimientos logísticos, las especificaciones de pruebas y los procedimientos para calcular puntajes en las pruebas. **Si bien la UMC ha publicado un marco de trabajo (2009) que explicita los objetivos y características de la evaluación²⁷, faltaría publicar un informe técnico más detallado de los procedimientos seguidos, análisis y resultados.** Tal información es útil, por un lado, para fortalecer la legitimidad de las evaluaciones, pero, por otro, para investigadores y otros actores que quieran usar la información en análisis particulares. Sugerimos destinar tiempo y recursos para este tipo de tareas desde el inicio de un ciclo de evaluación.

7. Sobre la institucionalidad de la evaluación. Si bien, como se ha indicado arriba, el CNE reconoce la legitimidad de la UMC como instancia evaluadora, nos preguntamos si corresponde que un ente evaluador se mantenga como una oficina dentro del Ministerio de Educación o que, más bien, sea parte de una institución pública, pero independiente. Mantenerla en el Ministerio tiene la ventaja de facilitar una mayor coordinación con otras instancias que son destinatarias principales de sus resultados y mensajes. Por otra parte, conseguir su autonomía del MED permitiría seguramente mayor flexibilidad en cuanto a la definición de los planes de evaluación y disponibilidad para la ejecución de sus recursos financieros, así como mayor independencia para interpretar los resultados. No queremos decir con esto que haya habido intromisión política reciente; de hecho, en la actualidad es prácticamente impensable que alguna autoridad política pudiera ocultar resultados, como ocurrió en los años 90, es prácticamente impensable. Nos referimos a señales muy concretas de independencia, como por ejemplo, que en los últimos años los resultados de las evaluaciones hayan sido presentados por el ministro o ministra de turno, lo cual no da una imagen de completa independencia. Un paralelo tal vez sea la presentación de cifras de pobreza, que en el Perú hace años no hace el Ministro de Economía, sino el Jefe del INEI. De hecho, en las discusiones sobre la ley del Sineace, se consideró que las tareas que hace actualmente la UMC podrían ser atribuidas al Ipeba, aunque esto finalmente no fue aprobado. Cuatro de los países con sistemas de evaluación más reconocidos (Brasil, Chile, Colombia y México) han formado institutos de evaluación adscritos al Estado, pero con algún nivel de autonomía respecto del Ministerio de Educación (ver detalles en anexos). Pasarle la UMC al Sineace tendría un beneficio adicional, y es que visto que las tareas de acreditación institucional, certificación de competencias profesionales (ambas a cargo del Sineace) y evaluación estandarizada del rendimiento (a cargo de la UMC, que es una dependencia del Ministerio de Educación) comparten los mismos principios, la labor de unos equipos podría fortalecer la de los otros. **El CNE considera que convendría retomar la discusión pública respecto de si la UMC debería seguir siendo una oficina del Ministerio de Educación o, en cambio, pasar a ser parte del Sineace. A corto plazo, sugerimos que se forme un Comité Externo de Evaluación que trabaje con la UMC en delinear las principales rutas de evaluación y que respalde, de ser el caso, los procedimientos técnicos empleados.** Este Comité se hace particularmente relevante, pues el CNE percibe que la evaluación estaría entrando en una cuarta etapa en el país, que podríamos denominar de *evaluaciones diversificadas*, para la cual convendría tener diversidad de opiniones y no solo las del Ministerio de Educación.

Finalmente, quisiéramos recordar el sentido principal de la evaluación, que es generar información válida y confiable que pueda ser utilizada como insumo para la mejora en la calidad y equidad del sistema educativo. Hemos visto con preocupación que, en ocasiones, se ha querido usar la

27 Ver http://www2.minedu.gob.pe/umc/ece/Marco_de_Trabajo_ECE.pdf.

evaluación para legitimar ciertos programas. Por ejemplo, recientemente se presentaron resultados de las instituciones educativas que participaban del programa de acompañamiento, que resultaron por encima de las “no acompañadas”. Tales resultados tendrían que ser sometidos a un análisis estadístico riguroso para verificar que los grupos de tratamiento (acompañadas) y control (no acompañadas) son similares (por ejemplo, en las características de los estudiantes y sus familias, y de las instituciones educativas), excepto en que unos participan en el programa y otros no. **El CNE sugiere no presentar resultados preliminares de impacto de programas, pues luego de un análisis riguroso se podrían demostrar errados.** Por lo demás, hay que ser cautos en la interpretación de los resultados, pues, como se sabe, los resultados educativos tienen una serie de factores explicativos, incluyendo los socioeconómicos, nutricionales y de salud, que interactúan con los propiamente educativos.

Reconocemos que las propuestas de evaluación presentadas arriba son complejas, pero el nivel de profesionalismo demostrado por la UMC nos hace pensar que con el apoyo requerido para las dos principales tareas sugeridas aquí, diversificación de las evaluaciones para incluir todos los aprendizajes fundamentales y diseño de estrategias para una mayor y mejor utilización de los resultados, sería posible realizarlas, siempre y cuando se contara con un incremento significativo de su presupuesto y personal especializado, acorde a la importancia de su labor. **El CNE recomienda por tanto incrementar los recursos financieros, humanos y logísticos disponibles para la UMC, de modo que se pueda iniciar reformas en las rutas sugeridas.**

Internacionalmente, las evaluaciones estandarizadas se van a incrementar. Ejemplos de esto son los planes de evaluaciones internacionales de la OCDE, Unesco e IEA mencionados antes. Relacionado con este incremento, el Banco Mundial ha planteado una nueva estrategia al 2020 basada en la meta *Learning for All* (Aprendizaje para todos) ²⁸ y el *Center for Global Development* en los Estados Unidos ha publicado recientemente un documento titulado *Escolaridad no es Educación: Usando las Evaluaciones para Cambiar las Políticas del No Aprendizaje*²⁹. Finalmente, la Unesco y el Instituto Brookings han formado un grupo internacional de trabajo sobre aprendizajes que se encuentra desarrollando tres documentos sobre cómo viabilizar las evaluaciones internacionales. Esto se relaciona con haber cumplido en muchos países con la meta educativa de acceso a educación (en el Perú, más del 90% de los niños y niñas en edad de asistir a la primaria lo hacen). La presión ahora es por demostrar que, más allá de asistir, los estudiantes aprenden en la escuela habilidades de diverso tipo que les permitirán, luego de culminar la secundaria, trabajar productivamente o continuar desarrollando sus habilidades a través de estudios superiores. Esta tendencia se va a sentir en el Perú de diversas formas, seguramente de manera más intensa que lo que se puede percibir actualmente. Siguiendo con el mensaje principal del presente boletín, debemos estar preparados no solo para evaluar, sino también para sacar el mejor provecho de los datos con los objetivos simultáneos de elevar la calidad y disminuir la inequidad educativa.

28 Ver <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTEDUCATION/0,,contentMDK:22474207~menuPK:282402~pagePK:210058~piPK:210062~theSitePK:282386,00.html>

29 Ver <http://international.cgdev.org/event/schooling-not-education-using-assessment-change-politics-non-learning>.

REFERENCIAS

Anderson, J. A. (2005). *Accountability in Education*. Paris: The International Institute for Educational Planning and the International Academy of Education.

Arregui, P. (Editora) (2006). *Sobre Estándares y Evaluaciones en América Latina*. Santiago de Chile: Preal.

Cueto, S., Guerrero, G., Zapata, M., y Freire, S. (2013). *Documento de Investigación Educación y aprendizajes*. Lima: Grade.

Ferrer, G. (2006a). *Sistemas de Evaluación de Aprendizajes en América Latina. Balance y Desafíos*. Santiago de Chile: Preal.

Ferrer, G. (2006b). *Estándares en educación. Implicancias para su aplicación en América Latina*. Santiago de Chile: Preal.

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Test-Based Accountability in Education* (p. 191). Santa Monica, California: Rand.

Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90(8-9), 1477–1503. doi:10.1016/j.jpubeco.2005.11.002

OECD (2003). *Literacy skills for tomorrow's world: further results from PISA 2000*. Paris: Unesco.

OECD (2010), PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I) <http://dx.doi.org/10.1787/9789264091450-en>

OECD. (2013). *Synergies for Better Learning. An international Perspective on Evaluation and Assessment. OECD Reviews of Evaluation and Assessments in Education*. Paris: OECD Publishing. doi:10.1787/9789264190658-en

Ravela, P. (2006). ¿Cómo aparecen los resultados de las evaluaciones educativas en la prensa? En Grupo de Trabajo sobre Estándares y Evaluación de Preal (Editores). *Sobre Estándares y Evaluaciones en América Latina*, (pp. 295-384). Santiago de Chile: Preal.

Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Rizo, F. M., Aylwin, M., et al. (2008). Las evaluaciones educativas que América Latina necesita. *Revista Iberoamericana de Evaluación Educativa*, 1(1), 45-63. Recuperado de http://www.rinace.net/riee/numeros/vol1-num1/art4_eng.html

UMC y Grade (2001). El Perú en el primer estudio internacional comparativo de la Unesco sobre lenguaje, matemática y factores asociados en el tercer y cuarto grado. *Boletín UMC* 9. Lima: Ministerio de Educación.

Unidad de Medición de la Calidad Educativa – UMC (2009). *Evaluación Censal de Estudiantes (ECE) Segundo grado de primaria, cuarto grado de primaria de IE EIB. Marco de trabajo*. Lima: Ministerio de Educación del Perú.

Unesco/Orealc (2008). Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer reporte de los resultados del Segundo Estudio Regional Comparativo y Explicativo. Santiago de Chile: Orealc/Unesco Santiago. Recuperado de <http://unesdoc.unesco.org/images/0016/001606/160660S.pdf>

Zambrano, G. (2002). *Las oportunidades de aprendizaje en lógico-matemática: un estudio para cuarto grado de primaria*. Lima: Ministerio de Educación del Perú.

ANEXOS

A1: Puntaje promedio en matemática según gestión, género y característica

Matemática											
		2008		2009		2010		2011		2012	
		Promedio	Desviación	Promedio	Desviación	Promedio	Desviación	Promedio	Desviación	Promedio	Desviación
Gestión	No Estatal	545,5	95,9	565,5	115,5	553,5	101,3	555,6	98,6	550,1	102,5
	Estatal	504,4	88,9	507,7	101,9	503,6	113,2	507,5	105,4	513,2	109,4
Género	Hombre	513,9	93,1	523,1	109,5	518,1	115,4	522,1	108,5	525,4	112,6
	Mujer	510,9	90,3	516,6	105,0	512,8	109,4	516,7	102,9	519,5	114,9
Característica	Multigrado	489,6	76,8	474,3	80,0	469,0	99,3	458,2	71,1	463,1	94,6
	Completa	521,8	95,1	533,1	111,2	529,1	112,7	533,6	108,4	535,4	107,5

Nota: Los puntajes usados están estandarizados con media 500 y desviación estándar 100
Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2007-2012.

A2: Puntaje promedio en comprensión lectora según gestión, género y característica

Comprensión lectora											
		2008		2009		2010		2011		2012	
		Promedio	Desviación	Promedio	Desviación	Promedio	Desviación	Promedio	Desviación	Promedio	Desviación
Gestión	No Estatal	561,2	86,6	572,2	86,9	591,2	81,6	588,2	84,3	592,9	86,7
	Estatal	489,3	80,0	511,1	77,5	517,2	99,6	517,4	92,3	523,3	91,7
Género	Hombre	499,0	84,2	519,1	81,4	530,4	98,7	529,6	93,0	536,6	94,5
	Mujer	507,4	87,8	528,7	84,9	539,6	101,9	540,4	97,6	545,0	96,1
Característica	Multigrado	463,7	67,5	479,1	62,6	468,7	85,3	459,5	63,5	471,5	83,0
	Completa	519,5	87,7	536,9	84,3	554,2	96,3	552,4	93,7	555,7	91,2

Nota: Los puntajes usados están estandarizados con media 500 y desviación estándar 100
Fuente: Muestra de Control, Evaluaciones Censales Estudiantiles 2007-2012.

A3: Descripción de sistemas nacionales de evaluación en países seleccionados³⁰

CHILE

La evaluación de la calidad de la educación en Chile está a cargo de la Agencia de la Calidad de la Educación, creada mediante la Ley de Aseguramiento de la Calidad (Ley N.º 20.529) del 2011.

30 La información presentada en este acápite ha sido copiada o adaptada de los respectivos portales.

Es independiente del Ministerio de Educación, de la Superintendencia de Educación y del Consejo Nacional de Educación, los otros organismos contemplados en esa ley y que, juntos, integran el denominado Sistema de Aseguramiento de la Calidad de la Educación. Las funciones de la Agencia de la Calidad de la Educación son, primero, **evaluar los logros de aprendizaje de los alumnos** en función a los estándares que elabora el Ministerio y aprueba el Consejo. Esto se logra a través del Sistema de Medición de la Calidad Educativa (Simce) que emplea instrumentos y procedimientos estandarizados, además de otros aspectos como la convivencia escolar. Asimismo, **clasifica los establecimientos según los resultados de aprendizaje y otros indicadores de calidad educativa**, para lo que tiene en cuenta las características de los alumnos que no dependen del establecimiento, como su nivel de vulnerabilidad, de forma que los colegios sean comparados con otros que se encuentren en una situación similar. Se procede así con la finalidad de focalizar el trabajo de la Agencia para que se concentre, con mucha más fuerza, en los establecimientos que tienen grandes deficiencias y los apoye para salir adelante. De acuerdo con su rendimiento, los establecimientos pueden clasificarse en Alto, Medio, Medio-bajo o Insuficiente. Al colegio que cuatro años consecutivos es clasificado como insuficiente se le revoca el reconocimiento.

La agencia también **evalúa el desempeño de los establecimientos educacionales, directores y sostenedores**. Para esto, realiza requerimientos de información, visitas evaluativas en el terreno y usa otros métodos, como grupos focales o entrevistas. Sobre esta base, elabora un informe público que determina fortalezas y debilidades, y da sugerencias para mejorar. Finalmente, también **proporciona información a la comunidad**, puesto que la clasificación de las escuelas es de público conocimiento, y también se entrega un informe escrito a los sostenedores y al equipo directivo de cada establecimiento. Además, los padres y apoderados tienen acceso a la situación del colegio. Todos estos datos están disponibles en la página web de la Agencia de Calidad de la Educación.

El portal web de la Agencia de la Calidad de la Educación es: <http://www.agenciaeducacion.cl/>

BOLIVIA

El Artículo 83 de la Ley de la Educación N.º 070 “Avelino Siñani – Elizardo Pérez”, del 20 de diciembre de 2010, creó el Observatorio Plurinacional de la Calidad Educativa (OPCE), una entidad pública descentralizada dependiente del Ministerio de Educación, pero con personería jurídica, con patrimonio propio y autonomía de gestión técnica, administrativa, financiera y legal en el ámbito nacional. La OPCE es regulada en su estructura y funcionamiento por el D.S. N.º 0832 de 30 de marzo de 2011. Las funciones del OPCE son las siguientes: realizar la medición, evaluación y acreditación de la calidad de la educación regular, alternativa, especial, y de la formación magisterial; informar a las instituciones del Estado y a la comunidad de los resultados con la finalidad de orientar la definición de las políticas educativas; y proponer acciones de intervención frente a los problemas identificados.

En lo relativo a los aprendizajes de los alumnos, los estudios realizados por el OPCE caracterizan la situación de los procesos de aprendizaje en comprensión lectora, razonamiento lógico matemático y conocimiento de la realidad plurinacional. Las evaluaciones se realizan a través de muestras representativas y están dirigidas a estudiantes de 5.º y 8.º de primaria, y estudiantes de 4.º de secundaria en comprensión lectora, realidad plurinacional y razonamiento lógico-matemático.

El portal web del OPCE es: <http://www.opce.gob.bo/>

BRASIL

El Instituto Nacional de Estudios Educativos e Investigación Anísio Teixeira (INEP: <http://www.inep.gov.br/>), órgano evaluador del Ministerio de Educación, está orientado a recolectar, procesar y analizar datos relativos a la educación. Sus funciones son las siguientes: organizar y mantener el sistema de información y estadística educativa; planificar, orientar, y coordinar los sistemas y proyectos de evaluación educativa; apoyar a los estados, distritos federales, y municipios en la implementación

de sistemas y proyectos de evaluación educativa; coordinar la evaluación de los cursos de graduación, y definir los parámetros, criterios y mecanismos para la realización de exámenes de acceso a la educación superior; promover la difusión de información sobre la evaluación de la educación básica y superior; y articularse con instituciones nacionales, extranjeras e internacionales, mediante acciones de cooperación institucional, técnica y financiera, bilateral y multilateral.

En el Brasil existen dos exámenes, el Sistema Nacional de Evaluación de la Educación Básica (SAEB) y la Prueba Brasil (Prova Brasil). El primero, desarrollado a finales del decenio de 1980, empezó a implementarse en 1990. Fue reestructurado en 1995 para permitir las comparaciones de año a año. Se implementa cada dos años, de forma muestral, en portugués y matemática, a estudiantes de 4.º y 8.º año de educación básica y 3.º de alta secundaria. Los resultados se devuelven por estado y por país. La evaluación Prova Brasil fue establecida en el 2005 con la finalidad de recoger información más detallada que la SAEB, puesto que brinda detalles no solo por estado, sino también por municipalidad y escuela participante. Los resultados son públicos y de libre acceso.

También se implementa Provinha Brasil (Pruebita Brasil), una evaluación diagnóstica que recoge información sobre el nivel de alfabetización de los estudiantes de segundo grado y orienta el trabajo docente. Se implementa en dos etapas, al inicio y al fin del año escolar.

En el 2007, el Ministerio de Educación implementó el Índice de Calidad de la Educación Básica (Índice de Desenvolvimento da Educação Básica, IDEB) con la finalidad de monitorear el progreso educativo en todas las escuelas, municipalidades, estados y regiones del Brasil. El IDEB se construye con los resultados del SAEB y de Prova Brasil, junto con datos de matrícula y de las tasas de repetición y graduación, para generar un índice que cubre casi la totalidad de las 175 000 escuelas primarias y secundarias del Brasil.

COLOMBIA

El Instituto Colombiano para la Evaluación de la Educación - ICFES (ver <http://www.icfes.gov.co/>) es una empresa estatal de carácter social del sector Educación Nacional. Es una entidad pública descentralizada de orden nacional, de naturaleza especial, con personería jurídica, autonomía administrativa y patrimonio propio, vinculada al Ministerio de Educación Nacional (MEN). Ofrece servicios de evaluación de la educación en todos sus niveles, apoya al MEN en la realización de los exámenes del Estado y en investigaciones sobre los factores que inciden en la calidad educativa para ofrecer información pertinente y oportuna a fin de contribuir al mejoramiento de la calidad de la educación.

En la educación básica, este instituto desarrolla la evaluación Saber, una evaluación nacional que se aplica periódicamente a estudiantes de educación básica de todo el país, con el fin de conocer el desarrollo de sus competencias básicas en lenguaje, matemáticas y ciencias. Asimismo, tiene a su cargo los exámenes de Estado de la educación media (SABER 11) y de la educación superior (SABER PRO). También ofrece la oportunidad para que personas mayores de 18 años validen su bachillerato y es requisito obligatorio para acceder a la educación superior.

La primera evaluación Saber se administró en el 2009. La segunda, en el 2012, incluyó a los estudiantes de tercer grado, en lenguaje y matemáticas. Los estudiantes de quinto y noveno grados rindieron evaluaciones de lenguaje, matemáticas, ciencias naturales y competencias ciudadanas.

ECUADOR

El año 1996 se constituyó el sistema de medición de logros académicos APRENDO, el cual recoge información sobre el desempeño estudiantil de los escolares de 3.º, 7.º y 10.º grados de educación básica en matemática y comunicación. Se llevó a cabo los años 1996, 1997, 1998, 2000 y 2007. La evaluación de 2007 tuvo un carácter muestral y estuvo acompañada de la administración de factores asociados para directores, docentes y padres de familia.

El 2009 en el marco de la política sexta del Plan Decenal de Educación³¹, el Ministerio de Educación implantó el Sistema Nacional de Evaluación y Rendición Social de Cuentas³² que evalúa cuatro componentes: la gestión del Ministerio y sus dependencias, el desempeño de los docentes, el desempeño de los estudiantes y el currículo nacional. Sus objetivos fundamentales son el monitoreo de la calidad de la educación que brinda el sistema educativo ecuatoriano, y la definición de políticas que permitan mejorar los procesos de enseñanza y aprendizaje. El sistema de evaluación se complementa con la rendición social de cuentas que consiste en la entrega de información transparente a la ciudadanía sobre los resultados de los procesos de evaluación efectuados. Hasta el momento, este sistema no se ha implementado y no ha producido reportes.

MÉXICO

En México, gran parte de las evaluaciones son responsabilidad del Instituto Nacional de Evaluación Educativa (INEE, ver <http://www.inee.edu.mx/>), que es un organismo del gobierno público federal de carácter descentralizado, no sectorizado, con personalidad jurídica y patrimonio propio. El INEE cuenta con autonomía técnica, operativa y de decisión para el ejercicio de sus funciones. Su objeto consiste en generar, aportar y difundir información y conocimiento que permitan a las autoridades federales y locales evaluar diferentes componentes, procesos o resultados del Sistema Educativo Nacional en sus respectivos ámbitos de competencia, a fin de tomar decisiones que contribuyan a mejorar su calidad.

Los programas, servicios y acciones del INEE están orientados a la evaluación de la educación obligatoria (preescolar, primaria y secundaria) y de la educación media superior, tanto pública como privada. Además, ha de contribuir a la evaluación de la formación inicial y continua de los docentes, la educación para adultos, la especial e indígena.

Se espera que el INEE contribuya a garantizar el derecho a la educación al aportar elementos de información y conocimiento que apoyen la toma de decisiones para mejorar la calidad de la educación en el país. También se espera que fomenten la equidad en la distribución de oportunidades educativas, con fundamento en lo establecido por la Constitución Política de los Estados Unidos Mexicanos y en la Ley General de Educación.

31 Se puede descargar de: http://educacion.gob.ec/wp-content/uploads/downloads/2013/03/K1_Plan_Estrategico1.pdf

32 El documento que presenta la fundamentación, objetivos, estructura y funciones de este sistema se puede descargar de: http://www.oei.es/pdfs/SIST_NAC_EVALUACION.pdf

Presidente

Jesús Herrero Gómez S. J.

Comité Directivo

Andrés Cardó Franco
Hugo Díaz Díaz
Santiago Cueto Caballero
Gustavo Yamada Fukusaki

Secretaria ejecutiva

Nanci Torrejón Muñante

Asesor de Presidencia

José Ignacio López Soria

Autor

Santiago Cueto Caballero

Equipo técnico

Gonzalo Cobo Gonzales



Av. De la Policía 577 Jesús María, Lima
Teléfonos: 261-4322 / 261-9522
www.cne.gob.pe / prensa@cne.gob.pe

Para enviar sus comentarios, escríbanos a:
prensa@cne.gob.pe

Publicación auspiciada por:

