

CIPPEC



políticas públicas
public policies

Programa de Educación
Área de Desarrollo Social

DOCUMENTO DE TRABAJO N° 161
NOVIEMBRE DE 2017

¿Qué países mejoraron la calidad educativa? América Latina en las evaluaciones de aprendizajes

AXEL RIVAS | MARTÍN SCASSO

Índice

Resumen ejecutivo.....	10
Agradecimientos	16
Introducción.....	17
1. América Latina en PISA 2015.....	19
1.1. Una mirada de conjunto.....	19
1.2. La comparación de los países	32
2. Los resultados de América Latina en el tiempo: problemas y alertas sobre su comparabilidad ...	39
2.1. La comparabilidad de los resultados 2015	39
2.2. ¿Por qué el cambio de metodología en PISA 2015 afecta tanto a América Latina?	51
2.3. Comentarios finales sobre el cambio de metodología	57
3. Una reflexión necesaria: ¿qué nos dicen los resultados de las pruebas PISA?	58
3.1. El recorte específico de PISA	58
3.2. La validez de PISA.....	62
4. Resultados medidos por evaluaciones de la calidad por países	69
4.1. Argentina.....	69
4.2. Brasil	81
4.3. Chile.....	91
4.4. Colombia	100
4.5. México.....	111
4.6. Perú.....	120
4.7. Uruguay.....	127
Conclusiones.....	134
Bibliografía	137
Anexo I Indicadores sociales y educativos de los países.....	140
Indicadores de cobertura	141
Indicadores de trayectoria	143
Indicadores de equidad.....	145
Indicadores de contexto	145
Indicadores de inversión educativa.....	147
Anexo II El caso de Argentina en la prueba PISA 2015	149
Introducción general.....	149
La exclusión de Argentina en PISA 2015	149

El dilema de los resultados de Argentina en PISA 2015.....	154
Análisis de la composición de la muestra de Argentina	156
Análisis de distribución de resultados por escuela.....	159
Conclusiones.....	170
Anexo III - Incidencia de los ítems NRI en el porcentaje de respuestas correctas de PISA 2006 – 2015	175
Anexo IV Interrogantes sobre la calidad de las muestras de PISA en algunos países	178

Índice de cuadros y gráficos

Recuadro 1. Acerca de PISA.....	19
Gráfico 1. Porcentaje promedio de estudiantes en cada nivel de desempeño en Matemática (M), Lectura (L) y Ciencias (C) en PISA 2015, por región.....	20
Gráfico 2. Porcentaje de población en edad de asistir a la educación secundaria que se encuentra fuera de la escuela, por región. Circa 2015	22
Gráfico 3. Coeficiente de Gini y brecha de resultados entre los cuartiles I y IV. Países seleccionados. 2015	23
Gráfico 4. PBI per cápita y resultados PISA (promedio de las 3 áreas), por región. 2015.....	24
Gráfico 5. Gasto educativo por alumno de nivel secundario y resultados PISA (promedio de las 3 áreas evaluadas), por región. Circa 2015	24
Gráfico 6. Índice del status económico, social y cultural y resultados PISA (promedio de las 3 áreas evaluadas), por región. 2015	26
Gráfico 7. Variación de la tasa neta de cobertura en el nivel secundario y del resultado de las evaluaciones PISA. Países seleccionados. 2000-2012.....	27
Gráfico 8. Puntaje promedio en PISA, por área. Resultados originales y reescalados. Promedio de 7 países seleccionados de América Latina. 2000-2015.....	28
Gráfico 9. Variación del puntaje entre PISA 2000 y 2012 (promedio lectura y matemática) y de la brecha entre resultados de los cuartiles I y IV, por región.....	29
Gráfico 10. Variación de puntaje PISA 2000 - 2012 (promedio lectura y matemática). Cuartil I, Cuartil IV y total de evaluados, por región.....	30
Gráfico 11. Resultados de las evaluaciones SERCE y TERCE, área de lectura, 3er y 6to grado. 2006 – 2013	31
Gráfico 12. Resultados de las evaluaciones SERCE y TERCE, área de matemática, 3er y 6to grado. 2006 – 2013	31
Gráfico 13. Resultados de las evaluaciones SERCE y TERCE, área de ciencias, 6to grado. 2013 y 2006 – 2013	32
Gráfico 14. Porcentaje promedio de estudiantes en cada nivel de desempeño en Matemática (M), Lectura (L) y Ciencias (C) en PISA. Países seleccionados. 2015	33

Gráfico 15. Puntaje PISA 2015 e intervalos de confianza, por área y promedio tres áreas. Países seleccionados	34
Gráfico 16. PBI per cápita y resultados PISA (promedio), países seleccionados. 2015	36
Gráfico 17. Índice del status económico, social y cultural y resultados PISA (promedio de las 3 áreas evaluadas). Países seleccionados. 2015	36
Gráfico 18. Gasto educativo por alumno de nivel secundario y resultados PISA (promedio de las 3 áreas evaluadas). Países seleccionados. Circa 2015	37
Gráfico 19. Resultado evaluaciones PISA 2015 (promedio 3 áreas) y tasa neta de escolarización en el nivel secundario. Países seleccionados. 2000-2015	38
Gráfico 20. Resultados PISA 2006 - 2015, puntajes originales y reescalados (promedio tres áreas), por región	42
Gráfico 21. Resultados PISA 2006 - 2015, puntajes originales y reescalados, por área. Promedio de países seleccionados de América Latina.....	43
Cuadro 1. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), por área. Países seleccionados de América Latina.....	44
Gráfico 22. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), por área. Países seleccionados de América Latina.....	45
Gráfico 23. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), promedio tres áreas. Países seleccionados de América Latina.....	49
Cuadro 2. Porcentaje promedio de <i>non-reached items</i> (NRI) en las evaluaciones PISA, por región. Años 2000 a 2015.....	53
Cuadro 3. Porcentaje promedio de <i>non-reached items</i> (NRI) en las evaluaciones PISA, por país. Años 2000 a 2015.....	55
Cuadro 4. Niveles de desempeño en lectura en PISA.....	59
Cuadro 5. Niveles de desempeño en matemática en PISA.....	60
Cuadro 6. Niveles de desempeño en ciencia en PISA.....	60
Recuadro 2. Los ítems no alcanzados y los aprendizajes: ¿es pertinente el cambio metodológico?.....	64
Gráfico 24. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012.....	70
Gráfico 25. Puntaje promedio en lectura, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012	71
Gráfico 26. Puntaje promedio en matemática, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012	72
Gráfico 27. Puntaje promedio en ciencias, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012	72
Recuadro 3. Las evaluaciones nacionales de la calidad educativa en Argentina.....	73
Gráfico 28. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Argentina. 2006-2013.....	74

Gráfico 29. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Argentina. 2006-2013.....	74
Gráfico 30. Resultados ONE 2013 - Aprender 2016. 5to/6to año. Argentina. 2013-2016	75
Gráfico 31. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2013.....	76
Gráfico 32. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2013.....	76
Gráfico 33. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2016.....	77
Gráfico 34. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2016.....	77
Gráfico 35. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015	81
Gráfico 36. Puntaje promedio en lectura, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015.....	82
Gráfico 37. Puntaje promedio en matemática, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015	82
Gráfico 38. Puntaje promedio en ciencias, resultados originales y reescalados. Nivel primario y secundario. Brasil y promedio de países seleccionados de América Latina. 2000-2015.....	83
Recuadro 4. Las evaluaciones de la calidad educativa en Brasil.....	84
Gráfico 39. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015.....	84
Gráfico 40. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario Brasil. 1999-2015	85
Gráfico 41. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015.....	86
Gráfico 42. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015.....	86
Gráfico 43. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Brasil. 2005-2015.	87
Gráfico 44. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Brasil. 2005-2015.....	88
Gráfico 45. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015	91
Gráfico 46. Puntaje promedio en lectura, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015.....	92
Gráfico 47. Puntaje promedio en matemática, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015	92
Gráfico 48. Puntaje promedio en ciencias, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015	93

Recuadro 5. Las evaluaciones nacionales de la calidad educativa en Chile	94
Gráfico 49. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Chile. 2001-2016	94
Gráfico 50. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Chile. 2001-2016	95
Gráfico 51. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Chile. 2005-2016	96
Gráfico 52. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Chile. 2005-2016	96
Gráfico 53. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015	100
Gráfico 54. Puntaje promedio en lectura, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015	101
Gráfico 55. Puntaje promedio en ciencias, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015	101
Gráfico 56. Puntaje promedio en matemática, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015	102
Recuadro 6. Las evaluaciones nacionales de la calidad educativa en Colombia	103
Gráfico 57. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016	103
Gráfico 58. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016	104
Gráfico 59. Resultados en ciencias en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016	104
Gráfico 60. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2005-2016	105
Gráfico 61. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2005-2016	105
Gráfico 62. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Colombia. 2006-2016	106
Gráfico 63. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Colombia. 2006-2016	107
Gráfico 64. Resultados en ciencias en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016	107
Gráfico 65. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015	111
Gráfico 66. Puntaje promedio en matemática, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015	112
Gráfico 67. Puntaje promedio en ciencias, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015.	112

Gráfico 68. Puntaje promedio en lectura, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015.	113
Recuadro 7. Las evaluaciones nacionales de la calidad educativa en México	114
Gráfico 69. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. México. 2006-2015.....	114
Gráfico 70. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. México. 2006-2015.....	115
Gráfico 71. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. México. 2006-2013	116
Gráfico 72. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. México. 2006-2013.....	116
Gráfico 73. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015	120
Gráfico 74. Puntaje promedio en lectura, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015.....	121
Gráfico 75. Puntaje promedio en matemática, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015	121
Gráfico 76. Puntaje promedio en ciencias, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015.....	122
Recuadro 8. Las evaluaciones nacionales de la calidad educativa en Perú.....	123
Gráfico 77. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Perú. 2007-2016	123
Gráfico 78. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Perú. 2007-2016.....	124
Gráfico 79. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015	127
Gráfico 80. Puntaje promedio en lectura, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015	128
Gráfico 81. Puntaje promedio en matemática, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015	129
Gráfico 82. Puntaje promedio en ciencias, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015	129
Recuadro 9. Las evaluaciones nacionales de la calidad educativa en Uruguay	130
Gráfico 83. Resultados SERCE- TERCE en lectura. Uruguay. 2006-2013.....	131
Gráfico 84. Resultados SERCE- TERCE en matemática. Uruguay. 2006-2013	131
Cuadro 7. Períodos presidenciales. Países seleccionados. 2000-2015	140
Gráfico 85. Tasa de analfabetismo en la población de 15 y más años.....	141
Gráfico 86. Tasa neta de escolarización en el nivel secundario. Países seleccionados. 2000 y 2012... ..	142

Gráfico 87. Porcentaje de población en edad de asistir a la educación secundaria (*) que se encuentra fuera de la escuela. Países seleccionados. Circa 2015	142
Gráfico 88. Tasa neta de escolarización en el nivel secundario por quintiles de ingreso. Países seleccionados. Circa 2000 y 2014.....	143
Gráfico 89. Tasa bruta de escolarización de la población de 3 a 5 años de edad. Países seleccionados. Circa 2000, 2007 y 2014.....	143
Gráfico 90. Evolución del porcentaje de alumnos repitentes, nivel primario (CINE 1). Países seleccionados. 2001 y 2015	144
Gráfico 91. Evolución del porcentaje de alumnos repitentes, nivel secundario básico (CINE 2). Países seleccionados. 2001 y 2015	144
Gráfico 92. Variación en los niveles de segregación escolar por nivel socioeconómico. Índice de Brecha por centiles (CGI)	145
Cuadro 8. Resumen de indicadores económicos y sociales. Países seleccionados, circa 2015 y variación circa 2000 - circa 2015	145
Gráfico 93. PBI y gasto público total como porcentaje del PBI. Países seleccionados. 2000-2015	146
Cuadro 9. Indicadores sociales complementarios. Países seleccionados, circa 2015 y variación circa 2000-circa 2015.....	146
Gráfico 94. Pobreza moderada y extrema (porcentajes) y PBI per cápita (dólares internacionales 2005 constantes). América Latina y el Caribe. 2000-2014	147
Gráfico 95. Gasto en educación como porcentaje del PBI y como porcentaje del gasto público total. Promedio de países seleccionados. 2000-2013.....	147
Gráfico 96. Evolución del gasto en educación frente al PBI. Países seleccionados. 2000-2014.....	148
Cuadro 10. Datos de cobertura de PISA 2015 en Argentina.....	151
Gráfico 97. Resultados PISA 2006 - 2015. Argentina, CABA y promedio de países de América Latina.....	155
Cuadro 11. Datos e indicadores seleccionados de PISA 2012 y 2015. Argentina.....	157
Gráfico 98. Porcentaje de escuelas según rango de resultados. PISA 2012 y 2015. Argentina.....	161
Gráfico 99. Porcentaje de <i>non-reached items</i> sobre el total de ítems aplicados, en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina.....	163
Gráfico 100. Porcentaje de respuestas correctas sobre el total de ítems válidos (descontando no respuesta y NRI), en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina	164
Cuadro 12. Porcentaje de ítems NRI y % de respuestas correctas sobre el total de ítems válidos (descontando no respuesta y NRI), por tipo de ítem. Total escuelas 2012 y 2015, y escuelas con muy bajos resultados en 2012. Argentina	166
Gráfico 101. Promedio índice nivel socioeconómico y cultural (ESCS) en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina.....	167
Gráfico 102. Porcentaje de escuelas según rango de resultados. PISA 2012. Argentina y países seleccionados	169

Cuadro 13. Porcentaje de ítems con respuestas correctas, calculadas según método A y B, por área y país. Años 2006-2015.....	176
Cuadro 14. Índices de cobertura 1,2 y 3 de PISA 2015. Países de América Latina	181
Cuadro 15. Población de 15 años, población objetivo de PISA e índice de cobertura 3. PISA 2015. Países de América Latina.....	182
Cuadro 16. Indicadores de cobertura vinculados a los estándares técnicos de calidad de PISA. PISA 2015	183
Cuadro 17. Índices de cobertura 4 y 5 de PISA 2015, y sus componentes. Países seleccionados de América Latina.....	184
Cuadro 18. Índice de cobertura 4 de PISA. Países de América Latina	185
Cuadro 19. Índices de cobertura 1 a 5 de PISA. Países seleccionados.....	186

Resumen ejecutivo

Este documento caracteriza la evolución de la calidad educativa medida por las evaluaciones internacionales y nacionales en América Latina en los años recientes. El enfoque que se propone toma como antecedente el libro “América Latina después de PISA. Lecciones aprendidas de la educación en siete países” (Rivas, 2015)¹, y se orienta a trascender la mirada basada en un indicador o una fuente única de datos, abriendo el análisis a la multiplicidad de información disponible sobre los aprendizajes y los contextos en los que se desarrollan.

El estudio se centró en la pregunta sobre la evolución de los resultados educativos medidos por las evaluaciones estandarizadas en América Latina durante las últimas dos décadas. Esto no implica asumir que la calidad educativa es igual a lo que se mide en las pruebas estandarizadas, que constituyen un recorte limitado. Estas pruebas permiten dar cuenta parcial del cumplimiento del derecho a la educación en términos de los aprendizajes que logran demostrar los alumnos en determinadas edades.

Conocer la evolución de cada país permite comprender mejor la situación del derecho a la educación en América Latina. Tomar estos resultados permite a su vez aprender de los países que han logrado mejorar en diversos indicadores. Esto implica un cuidadoso trabajo metodológico para no arribar a conclusiones apresuradas, tal como intenta mostrar este trabajo.

La publicación de los resultados de PISA 2015 de la OCDE en el mes de diciembre de 2016 constituye una nueva oportunidad para acceder a un caudal diverso y prolífero de información que permite evaluar los avances de América Latina en materia de calidad y equidad de los aprendizajes. Esta edición contó con la participación de los siete países que, desde hace varios ciclos atrás, implementan la prueba PISA en forma continua: Argentina², Brasil, Chile, Colombia, México, Perú y Uruguay). Se suma Costa Rica con su segunda aplicación, y la primera experiencia de República Dominicana. En este conjunto de países asiste aproximadamente el 80% de los estudiantes de los países de América Latina y el Caribe.

Los resultados muestran a la región con amplios desafíos en materia de aprendizajes, dando cuenta de la existencia de un conjunto mayoritario de adolescentes, especialmente de los sectores populares, que están en la escuela pero no logran adquirir aprendizajes para construir un proyecto autónomo de vida que les permita cambiar su destino social. Esto se combina con el todavía importante número de jóvenes que no asisten a la escuela secundaria o media en la región y que ni siquiera llegan a ser evaluados por las pruebas PISA.

Como señal positiva, una primera comparación directa simple de los resultados con el ciclo de PISA 2012, invita a situar a la región en un plano de mejora, al observarse un incremento generalizado en los desempeños de los estudiantes. Este escenario promisorio se ve afectado por una situación particular que caracteriza esta aplicación de PISA, que demanda mirar estos resultados desde otra óptica. Los cambios metodológicos implementados en PISA 2015 afectan la comparabilidad

¹ El libro “América Latina después de PISA. Lecciones aprendidas de la educación en siete países”, elaborado por el Programa de Educación de CIPPEC con el apoyo de Natura regional e Instituto Natura, presenta los resultados de una profunda indagación comparada de las políticas educativas implementadas en siete países durante el período 2000-2012, y su vinculación con hipótesis sobre la evolución de los distintos indicadores de la calidad educativa de los países. Véase: <http://mapeal.cippec.org/>

² Cabe destacar que la OECD definió la invalidación de los resultados de Argentina en la prueba PISA 2015, como consecuencia de errores detectados en la confección del marco de muestreo.

intertemporal en varios países de América Latina, y obligan a revisar la evolución de los desempeños desde las primeras aplicaciones.

PISA introdujo en 2015 cambios en el escalamiento de los resultados con el objetivo de mejorar la precisión de las mediciones y reducir los factores de incertidumbre. Si bien se mantuvieron las escalas y los ítems de anclaje que garantizan la comparabilidad, fueron introducidos algunas transformaciones vinculadas con el modo de implementación y el modelo de escalamiento de los resultados.

Para el conjunto de países evaluados, estas modificaciones inciden levemente en la comparabilidad entre años. Sin embargo, la mirada detallada sobre los países de América Latina pone de manifiesto que los resultados se ven sensiblemente afectados en su tendencia en el tiempo. Mejoras sustantivas que se observaban en los últimos ciclos de PISA en algunos países de la región se atenúan o desaparecen completamente al analizarlas con la nueva metodología. Lamentablemente, PISA puso a disposición los resultados reescalados hasta el año 2006, restringiendo las posibilidades de análisis del impacto de este cambio a los últimos nueve años de PISA, quedando fuera de esta comparación las pruebas de los años 2000 y 2003.

Para algunos países en particular, como Colombia, Perú y Uruguay, el escenario que perfilan los resultados con la nueva metodología implica un cambio importante en las tendencias de caracterización de la calidad de los aprendizajes.

Es necesario explicar por qué sólo los países de América Latina se han visto afectados por este cambio de metodología. La hipótesis que se propone sitúa la mirada en el tratamiento de los ítems no alcanzados (*non-reached items*) para el cálculo de los resultados. Con este término se hace referencia al conjunto de ítems que quedan sin responder al final del cuadernillo o de cada bloque correspondiente a un área, porque el estudiante no llega a completar la prueba o la abandona.

En las ediciones anteriores, los *non-reached items* se consideraban como respuestas incorrectas, mientras que en el 2015 fueron interpretados como ítems no administrados. Es decir, en los resultados de PISA 2000 a 2012, los estudiantes que dejaban la prueba con ítems incompletos al final se veían perjudicados en sus resultados, pero a partir del 2015 esto no ocurre.

Un análisis pormenorizado de las respuestas en todas las aplicaciones de PISA pone de manifiesto que la existencia de no respuesta al final de las evaluaciones es un fenómeno casi exclusivo de los países de América Latina. En las primeras aplicaciones de PISA en la región, cerca del 10% de los ítems se consideraban *non-reached*. Esta proporción fue disminuyendo en las aplicaciones posteriores. Por esta razón, el recálculo de los resultados de ciclos anteriores con la metodología nueva da como resultados puntajes superiores y, por ende, menor crecimiento respecto de los resultados actuales.

La imputación de estos ítems a respuestas incorrectas se sustenta en el supuesto de que pruebas incompletas expresan exclusivamente dificultades en el desempeño, asociadas a la falta de tiempo para completarla. Sin embargo, existen otros factores poco estudiados, vinculados a la disposición del estudiante y la escuela a responder con esfuerzo a las evaluaciones y al conocimiento de estos formatos³, que inciden en la existencia de pruebas incompletas.

³ Algunos autores vinculan estas características con el desarrollo de una “cultura” de la evaluación en los países participantes de PISA (Jodouin, 2014).

Es decir, la imputación de los *non-reached items* como respuesta errónea puede asociarse tanto con criterios vinculados con los conocimientos de los estudiantes, pero también con ciertas condiciones en las que se aplica la prueba. De esta manera, los cambios metodológicos de la edición 2015 dejan al desnudo dos aspectos críticos: la existencia de ciertas condiciones de preparación en el universo evaluado que inciden en los resultados, y la manera en que PISA presenta y comunica sus resultados.

El problema de comunicación de los resultados 2015 por parte de la OCDE refiere a la poca visibilidad con que fueron presentados estos cambios metodológicos. Afectaron la comparabilidad intertemporal de los resultados, especialmente en algunos países de América Latina, por lo que podrían haber sido aclarados en el cuerpo central de los informes de PISA y en los informes por países, alertando sobre las implicancias de la comparación directa.

El otro problema refiere a la manera en que la prueba PISA es implementada por los países. Por un lado, remite al modo en que las escuelas y los estudiantes se disponen a responder una evaluación que no mide sus aprendizajes de forma individual ni tiene consecuencias sobre ellos, sino que mide un sistema educativo completo. Esto implica la disposición de hacer el mayor esfuerzo para un fin relativamente abstracto: el mejor logro del sistema educativo nacional.

Un factor clave parece ser la preparación directa para realizar las pruebas PISA que la mayoría de los países de la región comenzó a implementar a medida que pasaron las sucesivas ediciones. Estas acciones se orientan a favorecer la captura de información de calidad, promoviendo mediante diferentes estrategias que las respuestas a la evaluación reflejen lo más fielmente posible las diferencias en los aprendizajes de los estudiantes, sin intervención de otros factores.

El efecto del cambio de metodología en el análisis de tendencias invita a reflexionar en torno a un interrogante clave: ¿en qué medida los resultados de América Latina mejoraron en los últimos 15 años por un incremento en la calidad de los aprendizajes, o como efecto de aplicaciones más cuidadosas? ¿Cambió el nivel promedio de esfuerzo de los alumnos al responder las pruebas PISA en sus sucesivas ediciones? Si cambió el nivel de esfuerzo de los alumnos al contestar la prueba, ¿cómo podemos saber si realmente mejoraron (o no empeoraron) sus aprendizajes o si solo mejoró su predisposición a responder la prueba con su máximo esfuerzo?

No hay forma de contestar esta pregunta con la investigación realizada. Pero obliga a establecer algunas reflexiones acerca de los alcances y las limitaciones de la prueba PISA para elaborar conclusiones sobre la calidad educativa en los países de la región. Éstas no se proponen poner en duda los cimientos sobre los que se sostiene PISA, ni a dar por tierra la calidad técnica de sus resultados. La comparación de los resultados entre países sigue siendo válida, especialmente si se observan varias ediciones de las pruebas PISA.

Otro aspecto positivo es que las comparaciones hacia adelante serán más fiables, tanto por el aumento del compromiso con la realización de la prueba, como por los cambios metodológicos introducidos por PISA en 2015. Pero la evolución por país a lo largo del tiempo debe ser analizada con recaudos en América Latina, evitando sacar conclusiones apresuradas sobre los cambios producidos de una edición a otra. La mirada debe ser complementada por el análisis de otras evaluaciones estandarizadas, nacionales e internacionales, que permitan triangular las tendencias observadas.

Por otro lado, es importante considerar que PISA no es una evidencia suficiente para realizar afirmaciones contundentes sobre la mejora de ciertos países, y menos aún para tomar lecciones de política de ellos. Recomendar políticas educativas requiere un balance de evidencias empíricas con estudios más profundos sobre los contextos culturales, las condiciones económicas y sociales y visiones de la educación en términos políticos, filosóficos, pedagógicos, curriculares y económicos.

Este marco general de abordaje lleva a profundizar la mirada sobre las tendencias específicas de cada país. El análisis cruzado entre resultados PISA con otros operativos internacionales, con evaluaciones nacionales y con datos de contexto arroja algunas observaciones relevantes:

- Perú muestra un camino de mejoramiento educativo notable en el siglo XXI, con mejoras de gran magnitud en todas las pruebas analizadas y grandes avances en la inclusión y la trayectoria de los alumnos, aunque aún manifiesta grandes deudas en el terreno de las desigualdades.
- Partiendo de una situación muy distinta, Chile parece haber mejorado en los años primeros años de implementación de PISA, aunque luego encuentra una etapa de estancamiento que lo mantiene por encima de los demás países analizados. La evolución en las pruebas nacionales, sin embargo, no es tan clara. Estos logros se dieron en un período de crecimiento del acceso y mejora de las trayectorias educativas de los alumnos, que se encuentran en la actualidad claramente por encima del resto de los países evaluados de la región.
- Brasil parece tener una marcada diferencia entre el nivel primario y medio. En el primer caso, el país mejora en las pruebas nacionales e internacionales. En el nivel medio, en cambio, la prueba PISA y las pruebas nacionales no muestran avances significativos. Estos resultados se enmarcan en un ciclo de gran expansión del acceso a la educación secundaria y de reducción de la repetición en todo el sistema.
- México también tiende a una estabilidad en la educación secundaria con mejoras en el nivel primario. Lo más notable es la mejora en matemática en el nivel primario. También México logra en este período un gran incremento de las tasas de acceso al nivel medio, desde puntos de partida muy bajos y todavía lejanos a los demás países estudiados.
- En Colombia, los resultados muestran estabilidad en el primario y secundario, tanto en las pruebas internacionales de PISA (con resultados comparables) y UNESCO (para el nivel primario) como en las pruebas nacionales, donde hay vaivenes difíciles de explicar. En cambio, se constata un gran avance del país en el aumento de las tasas de escolaridad y reducción de la sobreedad en el nivel secundario.
- En Uruguay, se observa un ciclo de estabilidad en los resultados en primaria y secundaria. Con el reescalamiento de los resultados de PISA no se observan cambios en el período 2006-15 y no existen pruebas nacionales para contrastar estos avances.

Estas tendencias generales no son lineales, muchas veces se observan comportamientos contrapuestos según áreas de enseñanza evaluadas o niveles educativos. Esto requiere estudios más profundos sobre las diferencias dentro de cada sistema educativo, tomando en cuenta la gran diversidad territorial y la descentralización de la gestión escolar que tienen casi todos los países analizados.

Argentina merece un tratamiento aparte, debido a que sus resultados fueron considerados no válidos en la edición de PISA 2015. El caso demandó un estudio en profundidad de este país, que se presenta en el documento como anexo.

Dos elementos distintivos de PISA 2015 dificultan la interpretación de los resultados de Argentina: por un lado, en la publicación del informe, la OCDE toma la decisión de no incluir a Argentina en el análisis de los resultados del operativo, porque incumple con el estándar 1.7, uno de los criterios

básicos vinculados a la calidad de la información⁴. La causa se identifica en errores en la confección del marco muestral.

Por otro lado, los resultados de Argentina en PISA 2015 (publicados en el informe PISA en los anexos) muestran un incremento excepcional respecto de 2012, de una magnitud que tiene escasos antecedentes en esta evaluación.

Como ambos fenómenos (problemas en la muestra y resultados excepcionales) se presentan juntos, esto tiende a establecer una equivalencia apresurada: los desempeños de Argentina 2015 fueron altos *porque* la muestra tuvo problemas. Pero no existe ninguna evidencia que permita establecer *a priori* esta relación.

Por lo tanto, el documento explora algunas hipótesis que permitan explicar el llamativo incremento de los resultados. Se analiza la composición de la muestra de Argentina con variables sociodemográficas y educativas, y se reconoce muy similar a la de PISA 2012. Esto descarta inicialmente la existencia de sesgos asociados a un marco muestral distorsionado. La distribución de los resultados entre estudiantes y entre escuelas tampoco muestra señales claras que podrían indicar intervenciones externas asociadas a una manipulación de la muestra.

Si bien no es posible avanzar en explicaciones, dado que los interrogantes no pueden ser resueltos con los datos disponibles, existen ciertos indicios analizados que abonarían, por sobre otras posibilidades exploradas, la hipótesis que explica los resultados de 2015 en función de una mejora en las condiciones de aplicación de PISA. En esta hipótesis, las acciones de sensibilización desarrolladas para esa aplicación habrían promovido mejores resultados, pero no en términos de “manipulación”, sino en la obtención de información de mayor calidad, que refleja mejor los aprendizajes efectivos.

En esta hipótesis, el “problema” no estaría en 2015, sino en las aplicaciones anteriores, donde no se generaron las condiciones adecuadas para que los estudiantes se comprometan con la evaluación, resultando en una significativa presencia en algunas escuelas de pruebas incompletas y/o resueltas con desinterés. Si bien no es concluyente, este análisis vuelve a remitir a los mismos ejes mencionados previamente: la necesidad de fortalecer las condiciones de aplicación para que los estudiantes respondan a las pruebas con voluntad de expresar de la mejor forma posible sus saberes.

El análisis de los resultados de PISA 2015 en América Latina, enmarcado en un diálogo entre la mirada metodológica y el análisis de diferentes fuentes de datos, tiene un valor vital para las gestiones educativas de los países, en la comprensión de su propia evolución en las evaluaciones de la calidad y su rol como autoridades de ejecución de las pruebas.

Se han logrado grandes avances en la calidad de la implementación de las pruebas, no sin esfuerzos ni linealmente. Pero todavía hay grandes desafíos para evitar la existencia de dispares dosis de preparación y presión en la ejecución de las pruebas. Debe lograrse una estabilidad que garantice aplicaciones eficaces y resultados comparables. De lo contrario los resultados pueden tener una dosis de incertidumbre que lleve a confusiones. Este desafío abarca a las pruebas internacionales y nacionales.

Comprender mejor la dinámica de los sistemas educativos es una pieza vital del proceso de mejora. Requiere datos rigurosos, comparables en el espacio y el tiempo; investigaciones más profundas que

⁴ Fuente: Capítulo 14 (*Data Adjudication*) del informe técnico PISA 2015 (2017).

miren las culturas, las prácticas y la historia; miradas sensatas que busquen la verdad dentro de la racionalidad científica, evitando los peligrosos sesgos ideológicos que recorren los debates educativos y políticos de cada país. Sólo por esa vía es posible tener más fuerzas para garantizar el derecho a la educación, expandido en la capacidad de los alumnos de usar los aprendizajes para definir su propio destino.

Agradecimientos

Este trabajo fue elaborado por el Programa de Educación de CIPPEC con el apoyo de Natura regional e Instituto Natura. El estudio fue dirigido por Axel Rivas. La sistematización de las estadísticas fue coordinada por Martín Scasso. La escritura del documento fue una tarea conjunta de ambos autores. Agradecemos la colaboración de Macarena Feijoó en la sistematización de documentos metodológicos, el apoyo de Gerardo Bortolotto en el procesamiento de datos y la lectura comentada de las distintas fases del trabajo por parte de Alejandra Cardini y Belén Sánchez.

El documento preliminar fue enviado a distintos especialistas internacionales para su consulta técnica. Agradecemos especialmente las lecturas y comentarios de Felipe Martínez Rizo, Andrés Peri, Mariano Palamidessi, Pedro Ravela, Augusto Hoszowski y Rosalía Cortés. También agradecemos el envío de información adicional de la OCDE para analizar la evolución en el tiempo de las pruebas PISA y la lectura comentada de Francesco Avvisati del análisis metodológico.

Por último, queremos agradecer especialmente a los centenares de especialistas técnicos de las áreas de evaluación educativa que se han creado y potenciado en los países de América Latina y en los organismos internacionales, incluyendo, claro está, a los equipos de las pruebas PISA. Detrás del trabajo que se presenta a continuación hay incontables horas de trabajo de especialistas técnicos, aplicadores de las pruebas y numerosos esfuerzos que buscan comprender mejor la situación educativa de los países.

Hechos estos agradecimientos, es importante mencionar que toda la responsabilidad sobre los hallazgos y conclusiones de este documento son de los autores.

Introducción

Las evaluaciones de la calidad educativa son una aproximación reciente para comprender mejor el funcionamiento de los sistemas educativos. Casi todos los países de América Latina crearon dispositivos nacionales de evaluación que permiten realizar un seguimiento de la evolución de los aprendizajes en el tiempo. Las pruebas PISA (*Programme for International Student Assessment*, según sus siglas en inglés) de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), iniciadas en 2000, abrieron la gran compuerta de la comparación internacional. Estas evaluaciones aportan una gran dosis de información paradójica: puede ampliar las esferas de comprensión de lo que ocurre en la educación o sesgar las miradas facilitando lecturas simplificadas. El objetivo de este documento es aportar claridad y síntesis para aprovechar la enorme riqueza de las evaluaciones de la calidad, advirtiendo los riesgos de los usos apresurados y excesivos.

El punto de partida de este documento es el libro “América Latina después de PISA. Lecciones aprendidas de la educación en siete países” (Rivas, 2015). En ese estudio se realizó una profunda indagación comparada de las políticas educativas implementadas en siete países durante el período 2000-2015: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. En un trabajo conjunto con socios locales, liderado por CIPPEC, se analizaron los contextos sociales, políticos y económicos y los resultados educativos medidos por diversos indicadores disponibles. Luego de una amplia sistematización comparada, se elaboraron hipótesis sobre la evolución de los distintos indicadores de la calidad educativa de los países y sus posibles relaciones con las políticas implementadas.

El disparador de ese trabajo fue la participación continuada en el tiempo de estos siete países en las pruebas PISA, cuya primera edición del año 2000 se repitió cada 3 años. PISA brindaba una enorme cantidad de información comparada sobre los sistemas educativos de los países, permitiendo utilizar evidencia empírica inédita para comprender los resultados educativos y sus contextos hasta la edición 2012. En este documento se busca continuar el análisis de los hallazgos de esa investigación original con la actualización de los resultados de la prueba PISA del año 2015 (publicados en diciembre de 2016) y de las evaluaciones nacionales más recientes.

Frente a ese estudio más profundo, la pregunta de investigación que guía este documento es mucho más acotada: ¿cuál fue la evolución de la calidad educativa medida por las evaluaciones internacionales y nacionales en América Latina en los años recientes?

Seguir esta evolución nos permitiría también releer las hipótesis originales del estudio acerca de los posibles ejes que permiten entender por qué y cómo las políticas educativas impactan en los aprendizajes de los alumnos a escala sistémica. En definitiva, queremos entender más sobre cómo mejoran los sistemas educativos y para ello es inevitable saber si mejoran, cuáles y cuándo. Pero sólo podremos comenzar a indagar las hipótesis sobre por qué mejoran los sistemas educativos si entendemos para qué sirven las evaluaciones de la calidad y hasta qué punto son válidas para comprender la evolución de los países.

La metodología del presente estudio se centró en analizar la evolución de los resultados de aprendizaje medidos por las pruebas PISA y por cada una de las evaluaciones de la calidad nacionales de los siete países estudiados. A su vez, se realizó una actualización de una variada serie de estadísticas educativas de cada país, con el objetivo de ampliar la mirada hacia dimensiones centrales de los sistemas educativos: la inclusión, la trayectoria de los alumnos y la equidad socioeducativa.

Lo que iba a ser un estudio corto que buscaba el poder de síntesis y actualización de la información de un trabajo más amplio previo, encontró en su proceso de sistematización de los resultados de la prueba PISA un hallazgo metodológico que requirió un tratamiento especial. Dado

que la prueba PISA 2015 tuvo cambios metodológicos frente a ediciones anteriores, en este trabajo se analizó en detalle el impacto de esas modificaciones en la comparabilidad de los resultados entre los países de América Latina.

El documento comienza presentando los resultados actualizados de las pruebas PISA 2015 y luego aborda los cambios metodológicos de la prueba para situar la evolución de cada país participante de América Latina entre 2000 y 2015. Dado el efecto magnético que provocan las evaluaciones de la calidad, especialmente PISA, se incluyó un tercer apartado sobre la validez que tienen estas pruebas para medir los sistemas educativos. Allí se destaca el valor que aporta la prueba PISA y las limitaciones que tiene cuando se utiliza para extraer conclusiones apresuradas.

El cuarto apartado profundiza el análisis en cada país, analizando la trayectoria de las distintas evaluaciones nacionales de la calidad educativa y comparando sus resultados y metodologías con las evaluaciones internacionales de la UNESCO (SERCE y TERCE) y de la OCDE (PISA). A esto se suma una lectura cuidadosa de los indicadores disponibles de acceso y trayectorias educativas de los alumnos en los siete países estudiados.

En las conclusiones se retoman las hipótesis del estudio original (Rivas, 2015) sobre qué países habían logrado mejorar la calidad educativa medida por las evaluaciones y se sintetiza su evolución actualizada en el tiempo. El estudio muestra la complejidad del panorama comparado de los sistemas educativos de la región y relativiza las conclusiones apresuradas sobre los “casos de éxito” o los “fracasos” de los países. Sin caer en una peligrosa evocación de la eliminación de los sistemas de evaluación, que constituyen un claro avance para medir el cumplimiento del derecho a la educación en la región, se propone adoptar una mirada cuidadosa de su validez para comparar los países a lo largo del tiempo.

1. América Latina en PISA 2015

1.1. Una mirada de conjunto

La prueba PISA es aplicada por la OCDE en las áreas de lengua, matemática y ciencias a una muestra representativa de los alumnos de 15 años de cada país participante. En 2015 participaron 72 países, 35 de la OCDE y 9 de América Latina (OCDE, 2016a). El **recuadro 1** analiza las características técnicas centrales, mientras los apartados 2 y 3 se centran en los aspectos metodológicos y epistemológicos de las pruebas PISA.

Recuadro 1. Acerca de PISA

PISA (Programa Internacional para la Evaluación de Estudiantes) es una prueba de aplicación trienal llevada a cabo por la OCDE (Organización para la Cooperación y el Desarrollo Económicos) para evaluar sistemas educativos a partir de los conocimientos y habilidades de los alumnos de 15 años, edad que coincide con el fin de la educación obligatoria en muchos países. Desde 2000 hasta hoy fueron evaluados más de 80 sistemas educativos.

Las pruebas PISA no están directamente asociadas con el currículum: evalúan la medida en la que los estudiantes pueden aplicar sus conocimientos a la resolución de problemas y están equipados con las herramientas necesarias para la participación social plena. Para ello, se los evalúa a través de preguntas de opción múltiple y de respuesta abierta referidas a textos o gráficos propios de situaciones de la vida real. A su vez, tanto estudiantes como directivos deben responder cuestionarios de contexto que brindan información sobre variables escolares y extraescolares, necesarias para la interpretación de los resultados. PISA prevé también la inclusión de evaluaciones o cuestionarios adicionales que son optativos para los países.

La evaluación se aplica cada tres años en lectura, matemática y ciencias. En cada aplicación se evalúa con mayor profundidad una de las áreas: en 2000 y 2009, fue lectura; en 2003 y 2012, matemática y en 2006 y 2015, ciencias. Los estudiantes tienen dos horas para contestar la prueba, que se construye según un diseño matricial. En conjunto, se cubren unos 390 minutos de ítems de evaluación, aunque cada estudiante responde solo una selección del total de ítems que suma 120 minutos.

Aproximadamente, 540 mil estudiantes de 72 países y economías fueron seleccionados para ser evaluados por PISA en 2015. Ellos representan a cerca de 29 millones de adolescentes de 15 años que asisten a la escuela.

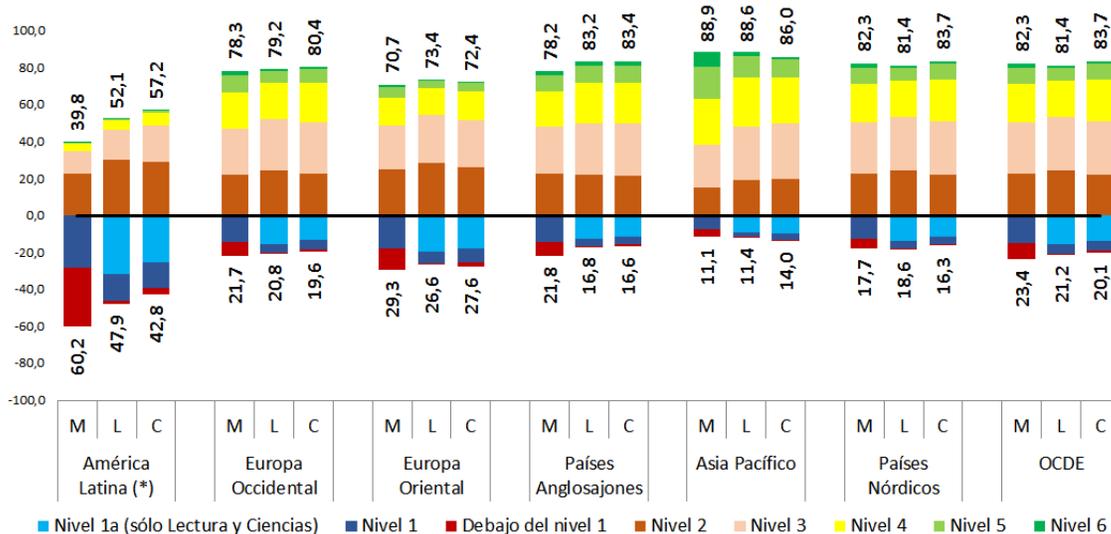
Desde la aplicación del año 2015 se generalizó el uso de la aplicación de la evaluación por computadora, 57 países aplicaron la prueba en esta modalidad. En América Latina, Argentina es el único país que mantuvo la aplicación en soporte papel.

Los resultados de las distintas aplicaciones de PISA (2000, 2003, 2006, 2009, 2012 y 2015) pueden compararse desde la aplicación en la que cada área fue evaluada en profundidad. Así, los resultados de lectura son comparables a partir de 2000, los de matemática son comparables a partir de 2003 y los de ciencias solo pueden compararse desde 2006.

Los resultados de PISA son expresados en una escala con promedio 500 y desviación estándar 100, construida sobre la base de metodologías alineadas con la Teoría de Respuesta al Ítem (TRI). Para cada área se definen 7 niveles de desempeño, asociados con determinados intervalos de puntaje. El nivel 2 de desempeño es considerado el umbral de competencias básicas para cada área evaluada.

La prueba PISA permite situar a los países de América Latina en sus logros de aprendizaje en comprensión lectora, matemática y ciencias. El **gráfico 1** sitúa a la región en su conjunto, comparada con el resto de los países participantes a partir de su agrupamiento en regiones específicas que permiten dimensionar distintos rasgos económicos, sociales y culturales comunes, así como también ciertas tradiciones pedagógicas. Allí se observa que los países de América Latina obtuvieron resultados de aprendizajes más bajos que los de otras regiones cuyos países que participan en PISA.

Gráfico 1. Porcentaje promedio de estudiantes en cada nivel de desempeño en PISA por región. Matemática (M), Lectura (L) y Ciencias (C), por región. 2015⁵



Referencias: M: matemática; L: lectura; C: ciencias.

Notas: el gráfico fue construido alineando los países en la frontera entre los niveles 2 y 1.

Las regiones están compuestas de la siguiente manera: América Latina (*): Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Austria, Bélgica, Francia, Alemania, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Albania, Bulgaria, Croacia, Rep. Checa, Estonia, Hungría, Kazajistán, Letonia, Lituania, Montenegro, Polonia, Rumania, Rusia, Rep. Eslovaca y Eslovenia. Países Anglosajones: Australia, Canadá, Nueva Zelanda, Reino Unido y EE.UU. Asia Pacífico: Taipei, Hong-Kong China, Japón, Corea, Macao-China, Shanghái-China y Singapur. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

(*) Los resultados de Argentina no fueron incluidos en el cálculo de América Latina debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.

Fuente: CIPPEC, sobre la base de OCDE, PISA 2015.

⁵ Debido a que el enfoque del documento se orienta a analizar y problematizar las tendencias en el tiempo de los desempeños en PISA, se decide excluir a Rep. Dominicana y Costa Rica de los análisis comparativos entre regiones. En el primer caso, el país sólo posee una participación en PISA 2015, y Costa Rica se incorpora a PISA en 2012. No se incluye a Argentina entre los países de América Latina dado que PISA excluyó al país por motivos metodológicos vinculados con el marco muestral de alumnos. Esta situación se analiza en detalle en el Anexo 3.

La proporción de alumnos por debajo del nivel 2 es muy alta: 60,2% en matemática, 47,9% en lengua y 42,8% en ciencias. En el otro extremo la proporción de alumnos es muy baja: apenas el 0,09% de los alumnos en ciencias, el 0,02% en lengua y el 0,08% en matemática alcanzó el nivel 6 en 2015, considerado el más alto de la escala. En comparación, los países de la OCDE tenían un 1,4% de alumnos en nivel 6 de ciencias, 1,2% en lectura y 2,1% en matemática.

La combinación de un bajo porcentaje de alumnos en el nivel 6 y de una proporción muy alta de alumnos por debajo de nivel 2 muestra un doble problema. El primer valor remite a la relativa ausencia de alumnos que a los 15 años de edad tengan niveles muy altos de formación para realizar futuros trabajos especializados y de mucha demanda cognitiva. El segundo y principal problema es la situación de vulneración de derechos de amplias capas de la población que no adquieren conocimientos fundamentales, entendidos como capacidades para resolver problemas matemáticos, comprender textos y usar el razonamiento científico.

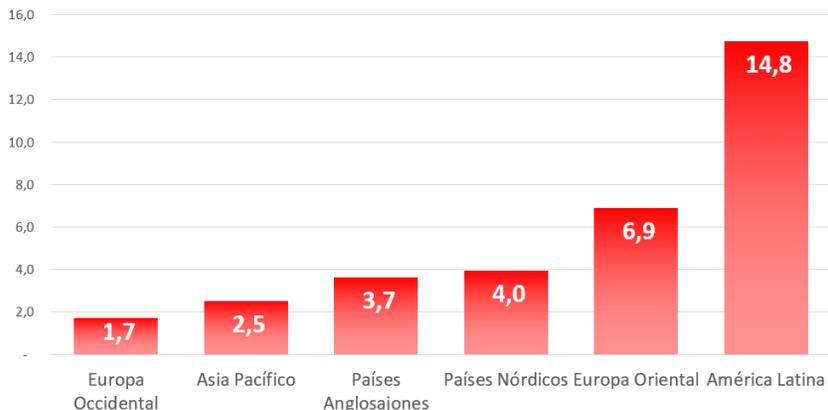
Esto no puede ser naturalizado: se trata de las amplias mayorías de América Latina, especialmente los sectores populares, que están en la escuela pero no logran adquirir aprendizajes para construir un proyecto autónomo de vida que les permita cambiar su destino social. El problema se combina con el todavía importante número de jóvenes que no asisten a la escuela secundaria o media en la región y que ni siquiera fueron evaluados por las pruebas PISA. Se estima que son 10,4 millones los adolescentes en edad de asistir a la educación secundaria que se encuentran fuera de la escuela en América Latina y el Caribe. De ellos, 7,6 millones habitan en alguno de los siete países bajo análisis en este documento, y representan el 14,8% de la población total⁶.

El **gráfico 2** permite identificar la incidencia de la exclusión educativa en los países analizados, en relación con las otras regiones participantes de PISA, y dimensionar el tamaño de los desafíos pendientes de inclusión.

Lamentablemente no se conoce la situación educativa de esta población excluida, que no es evaluada. En cambio, sí es posible identificar los desafíos de aprendizaje de aquellos que, habiendo logrado permanecer en la escuela, se encuentran en situación de desventaja.

⁶ En el anexo I se presentan los datos para cada uno de los países de América Latina participantes de PISA

Gráfico 2. Porcentaje de población en edad de asistir a la educación secundaria (*) que se encuentra fuera de la escuela, por región. Cerca 2015()**



(*) El cálculo considera a la población en edad teórica de asistir a la educación secundaria que no se encuentra asistiendo al nivel secundario ni primario. Se toma como referencia al rango de edad que corresponde al nivel secundario según cada estructura de niveles.

(**) Los datos corresponden a 2015 para todos los países excepto Argentina, Uruguay, Japón, Francia, Grecia, Luxemburgo, Kazajistán, Polonia, Eslovenia, Australia, Reino Unido y Estados Unidos (2014); Islandia (2013); México (2012) y Irlanda (2010)

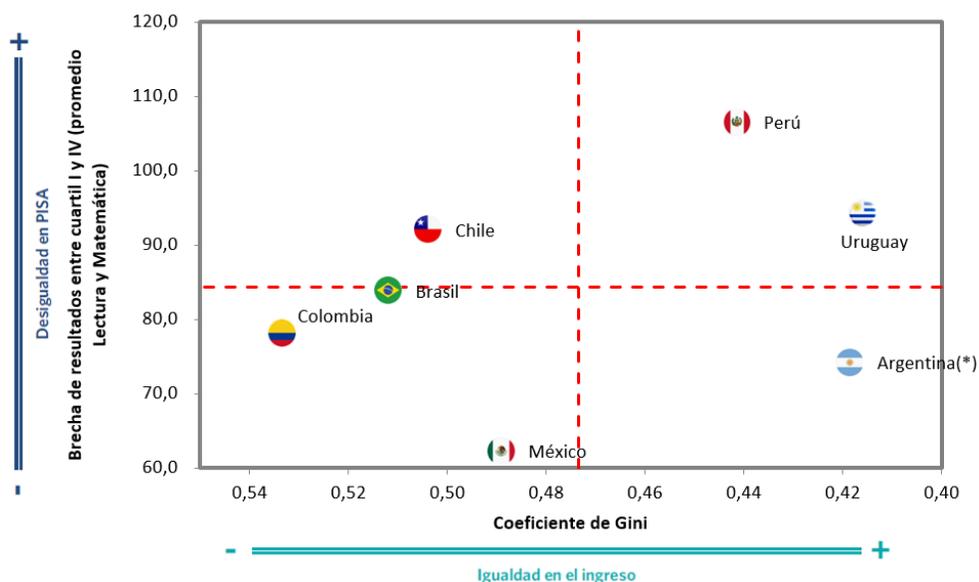
Las regiones están compuestas de la siguiente manera: América Latina: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Bélgica, Francia, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Albania, Bulgaria, Estonia, Hungría, Kazajistán, Letonia, Lituania, Polonia, Rumania y Eslovenia. Países Anglosajones: Australia, Nueva Zelanda, Reino Unido y EE.UU. Asia Pacífico: Hong-Kong China, Macao-China, Japón y Corea. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

Fuente: CIPPEC, sobre la base de UIS data centre

Los países de la región muestran amplias brechas en sus resultados entre los estudiantes más ricos y más pobres. En promedio, hay una diferencia de 85 puntos en los resultados entre ambos grupos. Esta distancia equivale a la diferencia de desempeños entre América Latina y los países nórdicos. Las inequidades se hacen más presentes en algunos países, y no parece mostrar una relación clara con los niveles de desigualdad social (aproximados a través del índice de Gini) o los niveles de cobertura.

Por ejemplo, México es el país con mayor grado de equidad en PISA, pero a la vez muestra los mayores niveles de exclusión. Chile y Uruguay muestran diferencias grandes en PISA entre ricos y pobres, con diferentes niveles de cobertura y de equidad social. Cada una de estas variables contribuye a la construcción de un escenario donde la desigualdad educativa debe leerse en contexto.

Gráfico 3. Coeficiente de Gini y brecha de resultados entre los cuartiles I y IV, países seleccionados. 2015



(*) Los resultados de Argentina fueron incluidos a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Los cuartiles se calcularon en función del Índice del status económico, social y cultural de la población, construido por OCDE a través de la información recolectada en los cuestionarios de PISA.

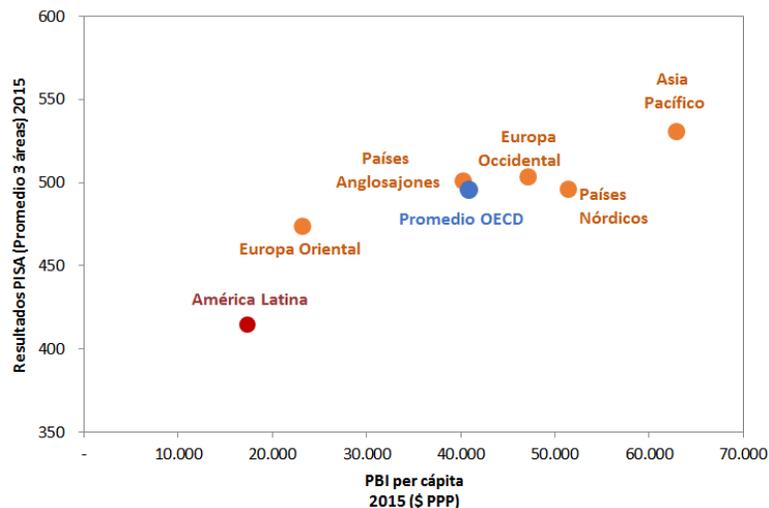
Fuente: CIPPEC, sobre la base de SEDLAC y OCDE, Bases de datos PISA 2015.

Los resultados en contexto

En referencia al contexto, los bajos resultados de aprendizaje de los países de América Latina en su conjunto parecen reflejar su situación económica y social. Los **gráficos 4, 5 y 6** muestran la correlación entre los resultados de la prueba PISA 2015 (expresados en un puntaje promedio de las tres áreas evaluadas⁷) y el PBI per cápita, la inversión por alumno del nivel secundario y el índice de status económico, social y cultural de los alumnos. En los tres casos se puede apreciar una clara correlación, que evidencia la fuerte correspondencia que existe entre estas tres variables y el nivel de aprendizaje de los alumnos.

⁷ El uso del promedio de las tres áreas como representación del desempeño se utiliza aquí con la exclusiva finalidad de construir una imagen simplificada y resumida de cada área en la comparación entre y como tendencias en el tiempo. No debe ser utilizado ni interpretado con otro sentido. Los niveles de desempeño y la interpretación de las escalas son diferentes en cada dominio evaluado por PISA, y cada uno de ellos está estandarizado de acuerdo a una población de referencia distinta. Por estas razones, las escalas no son comparables entre diferentes áreas. En este sentido, los puntajes promedio en las tres áreas representan únicamente la media aritmética de tres valores que no son comparables entre sí.

Gráfico 4. PBI per cápita y resultados PISA (promedio de las 3 áreas), por región. 2015



El PBI per cápita está expresado en dólares internacionales corrientes, calculados en base a la paridad del poder adquisitivo (PPA).

Las regiones están compuestas de la siguiente manera: América Latina (*): Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Austria, Bélgica, Francia, Alemania, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Albania, Bulgaria, Croacia, Rep. Checa, Estonia, Hungría, Kazajistán, Letonia, Lituania, Montenegro, Polonia, Rumania, Rusia, Rep. Eslovaca y Eslovenia. Países Anglosajones: Australia, Canadá, Nueva Zelanda, Reino Unido y EE.UU. Asia Pacífico: Hong-Kong China, Japón, Corea, Macao-China y Singapur.

Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

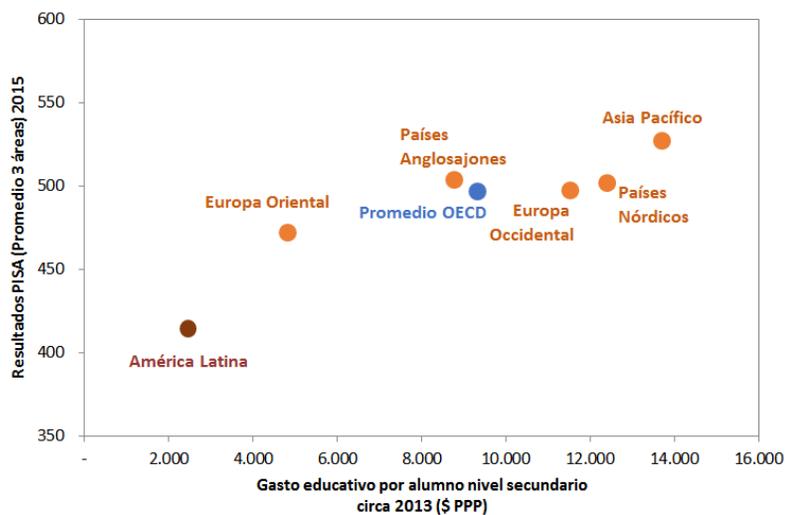
(*) Los resultados de Argentina no fueron incluidos debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.

Fuente: CIPPEC, sobre la base de OCDE-PISA 2015 y WEO-IMF.

América Latina es una región con economías en desarrollo más débiles que todas las demás regiones participantes en PISA (a través de sus países), con un promedio de casi 20 mil dólares anuales per cápita frente a los casi 50 mil de Europa Occidental. La inversión por alumno refleja esta desigualdad económica: en América Latina se invierte U\$S 2.460 por año por alumno, frente a los U\$S 9.335 de la OCDE o los U\$S 12.396 de los países nórdicos⁸. En la propia medición del contexto social de los alumnos que toman la prueba PISA se observa algo similar: América Latina tiene alumnos mucho más pobres y vulnerables que el resto de los países agrupados en regiones que participan de la prueba.

⁸ El gasto educativo por alumno de nivel secundario refiere al promedio del gasto público general (corriente, de capital y transferencias) en el nivel secundario, por alumno, expresado en valor nominal de dólares calculados en paridad de poder adquisitivo (PPA). Fuente: CIPPEC, en base a datos de WEO-IMF.

Gráfico 5. Gasto educativo por alumno de nivel secundario y resultados PISA (promedio de las 3 áreas evaluadas), por región. Circa 2015



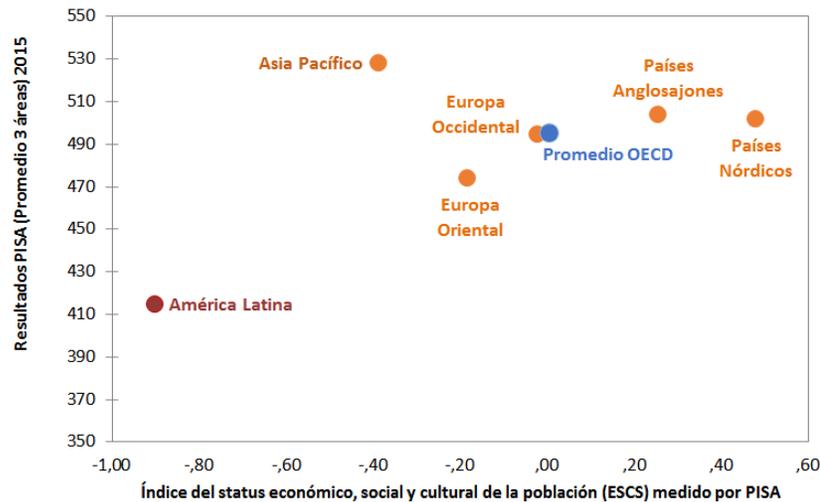
Nota: el gasto educativo por alumno de nivel secundario refiere al promedio del gasto público general (corriente, de capital y transferencias) en el nivel secundario, por alumno, expresado en valor nominal de dólares calculados en paridad de poder adquisitivo (PPA). Los años de referencia de los datos de gasto público por alumno son los últimos disponibles: 2015 para Perú, Hong Kong-China, Colombia y Kazajistán; 2014 para Chile, Japón, Nueva Zelanda y Reino Unido; 2013 para Albania, Australia, Austria, Brasil, Bulgaria, Macao-China, Rep. Checa, Dinamarca, Estonia, Francia, Alemania, Hungría, Islandia, Italia, Letonia, Lituania, Luxemburgo, México, Países Bajos, Polonia, Portugal, Corea, Rumania, Eslovaquia, Eslovenia, España, Estados Unidos y Suecia; 2012 para Irlanda y Suiza; 2011 para Canadá, Finlandia y Noruega.

Al no disponer de información actualizada, en Uruguay se utilizó el dato de gasto por alumno de 2006. Las regiones están compuestas de la siguiente manera: América Latina (*): Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Austria, Francia, Alemania, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Albania, Bulgaria, Rep. Checa, Estonia, Hungría, Kazajistán, Letonia, Lituania, Polonia, Rumania, Rep. Eslovaca y Eslovenia. Países Anglosajones: Australia, Canadá, Nueva Zelanda, Reino Unido y EE.UU. Asia Pacífico: Hong-Kong China, Japón, Corea y Macao-China. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

(*) Los resultados de Argentina no fueron incluidos debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.

Fuentes: CIPPEC, sobre la base de OCDE-PISA 2015 y UIS-UNESCO.

Gráfico 6. Índice del status económico, social y cultural y resultados PISA (promedio de las 3 áreas evaluadas), por región. 2015



Las regiones están compuestas de la siguiente manera:

América Latina (*): Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Austria, Bélgica, Francia, Alemania, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Albania, Bulgaria, Croacia, Rep. Checa, Estonia, Hungría, Kazajistán, Letonia, Lituania, Montenegro, Polonia, Rumania, Rusia, Rep. Eslovaca y Eslovenia. Países Anglosajones: Australia, Canadá, Nueva Zelanda, Reino Unido y EE.UU. Asia Pacífico: Hong-Kong China, Japón, Corea, Macao-China y Singapur. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

(*) Los resultados de Argentina no fueron incluidos debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.

Fuente: CIPPEC, sobre la base de OCDE-PISA 2015

Estos datos muestran lo que diversos estudios han analizado previamente en cuanto a la fuerte relación entre el contexto socioeconómico y los logros de aprendizaje (Sirin 2005). Todo esto indica que revertir los bajos resultados en PISA es una batalla compleja, que implica combinar la redistribución de la riqueza, el desarrollo económico sostenible, el aumento del financiamiento educativo y la mejora de la calidad y la capacidad de impacto de las políticas educativas.

Los resultados en el tiempo

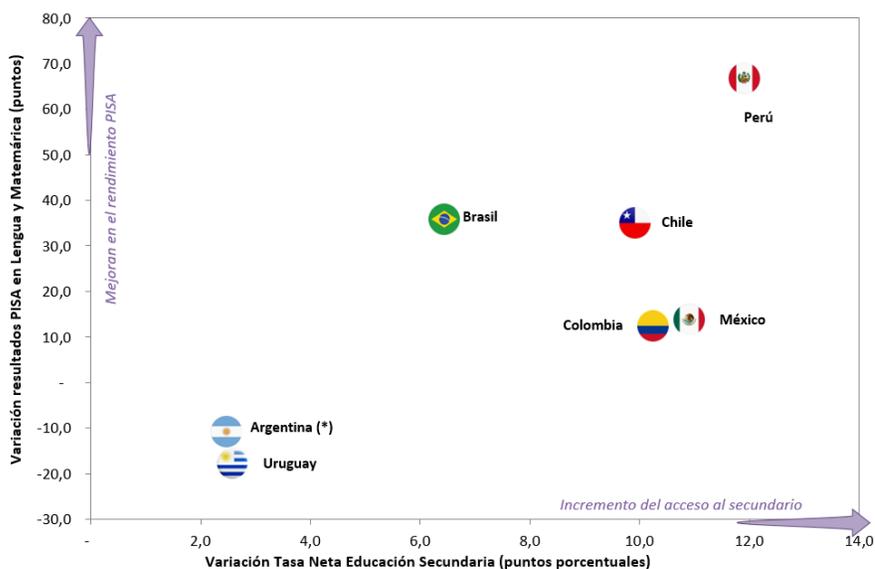
Pese a este difícil escenario que signa a la región más desigual del planeta, en los últimos 15 años América Latina mejoró tanto su acceso como su calidad educativa, medida por las propias pruebas PISA. Esto ya había sido demostrado en el estudio que antecede a esta publicación para el período 2000-2012 (Rivas 2015). La aparición de los resultados de la prueba PISA 2015 abre más interrogantes que continuidades en esta serie histórica. El motivo central de estos interrogantes es un cambio metodológico que afectó la comparabilidad de los resultados en la región, analizado en detalle en el siguiente apartado. La situación particular de Argentina, con un problema en su marco muestral también suma confusión a las posibles lecturas de la evolución en el tiempo.

No podemos afirmar con certeza cuál ha sido la evolución en el tiempo de los países de América Latina en PISA luego de 2012. Todo indica que, con las correcciones metodológicas, la región no ha seguido mejorando sino que ha encontrado mayormente un estancamiento que ya se advertía en la

prueba de 2012 frente a 2009. Quedan muchas preguntas abiertas sobre la mejora de varios países en el ciclo 2000-2009, por la falta de información comparable de los años 2000 y 2003.

Sin embargo, cabe destacar que esta mejora y el estancamiento posterior se dieron en un marco de gran expansión del acceso a la educación. América Latina logró mejorar la inclusión, la equidad y la calidad educativa en los primeros 15 años del siglo XXI. Como se advierte en el **gráfico 7**, la región mejoró sus indicadores de calidad medidos por la prueba PISA y las tasas de escolarización en el nivel secundario, que pasaron de 71,4% en 2000 a 79,2% en 2012 (y 81,5% en 2014, promedio simple) en los siete países estudiados en profundidad.

Gráfico 7. Variación de la tasa neta de cobertura en el nivel secundario y del resultado de las evaluaciones PISA, países seleccionados. 2000-2012



Nota: las tasas netas de educación secundaria circa 2000 y 2012 se tomaron de SITEAL, IPE-UNESCO y OEI (Argentina, Brasil México y Uruguay), SEDLAC-Banco Mundial (Chile, Colombia), y Estadística de la Calidad Educativa - ESCALE (Perú).

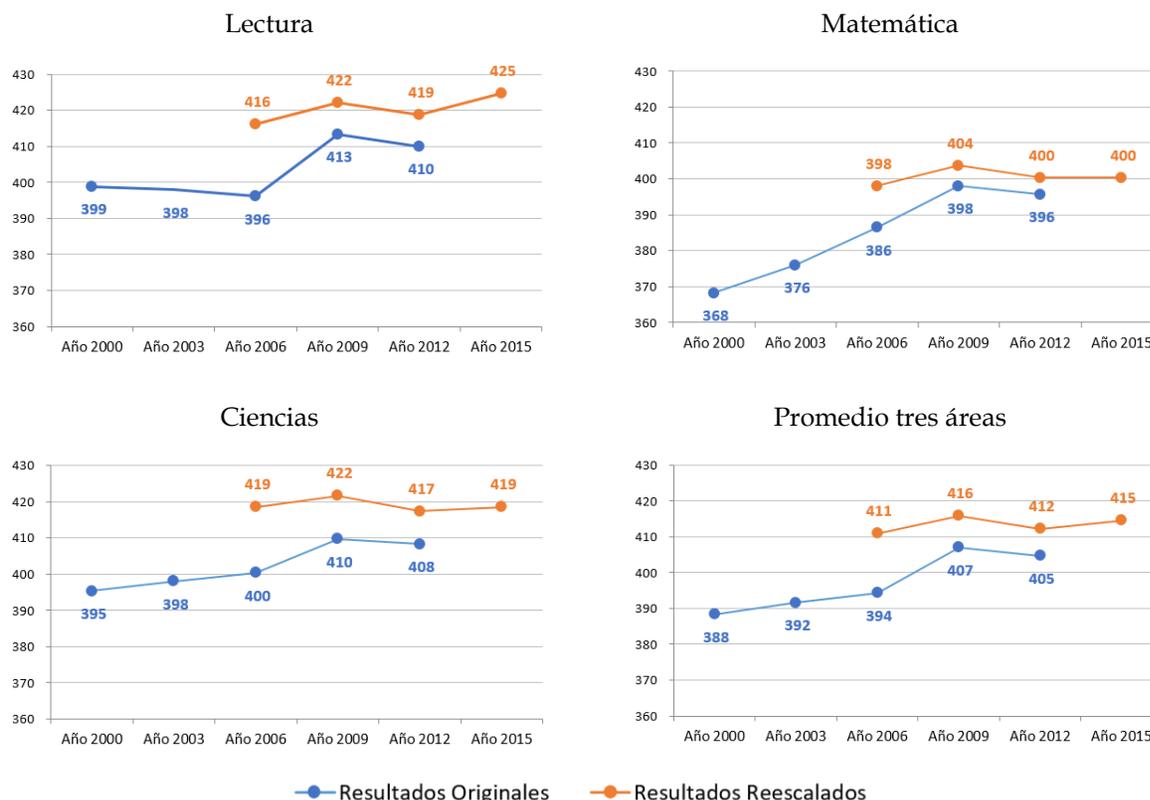
Fuente: CIPPEC, sobre la base de OCDE - PISA 2000 y 2015, SEDLAC, SITEAL, ESCALE (Perú).

En PISA los siete países de América Latina mejoraron en promedio 17 puntos entre 2000 y 2012 (según los resultados publicados en cada aplicación) y 4 entre 2006 y 2015⁹ (con el ejercicio de corrección metodológica y sin considerar a Argentina). Tomando como referencia los años en los cuales cada área evaluada es foco de PISA, se observa un aumento de 14 puntos en Lectura entre 2000 y 2009, de 20 puntos en Matemática entre 2003 y 2012 (con los resultados originales de cada aplicación) y resultados idénticos en ciencias entre 2006 y 2015 (con los resultados comparables, reescalados teniendo en cuenta los cambios metodológicos)¹⁰.

⁹ Para calcular los promedios se consideraron los primeros años de aplicación en los países que no participaron en el 2000 (Uruguay 2003 y Colombia 2006), y la proyección lineal simple en los países que no tienen aplicaciones intermedias (Perú 2006, en base a Perú 2000 - 2009).

¹⁰ Cabe aclarar que PISA centra cada edición de la prueba en un área de aprendizajes. Esto permite una comparación más rigurosa en términos metodológicos, especialmente desde que se inició la prueba. Es decir que se considera que el año 2000 es el de referencia para comparar lengua, el 2003 para matemática y el 2006 para Ciencia. La comparación de esos años con

Gráfico 8. Puntaje promedio en PISA, por área. Resultados originales y reescalados (*). Promedio de 7 países seleccionados de América Latina. 2000-2015



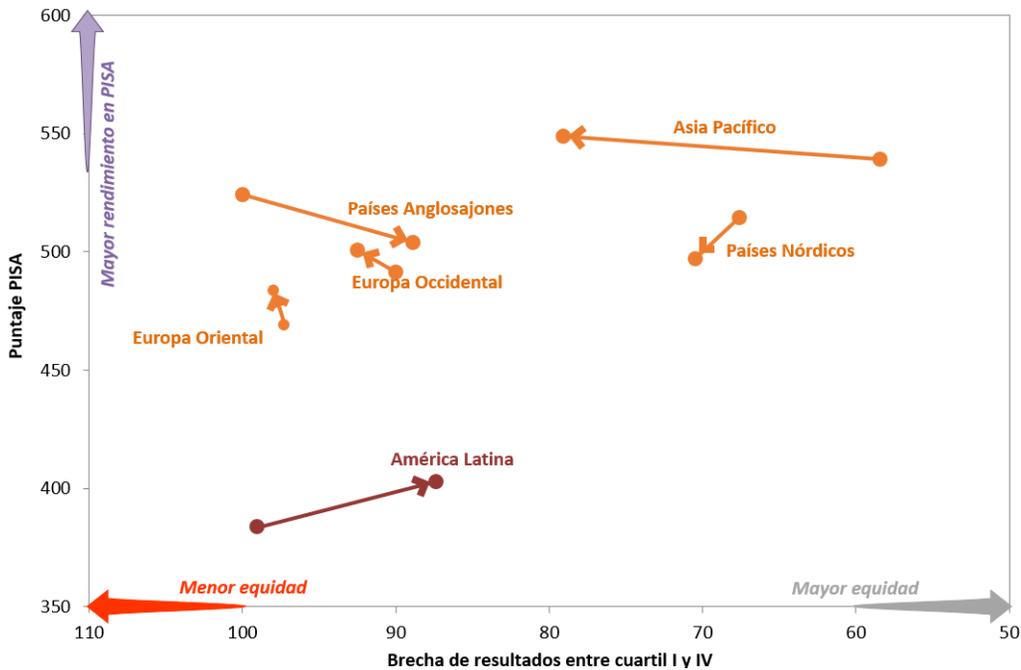
(*) Nota: Debido a los cambios metodológicos implementados en la edición 2015 de la prueba PISA, y analizados en el capítulo 2 de este documento, no es posible comparar directamente los resultados de 2015 con los de años anteriores. Para ello, deben utilizarse los desempeños de años anteriores reescalados por PISA con la nueva metodología, cuyos resultados solo fueron publicados hasta el año 2006. Por esta razón, se presentan dos series: la graficada en color naranja representa los resultados con la metodología de PISA 2015 (disponible hasta 2006), y la azul que incluye los resultados previos al cambio metodológico. Fuente: CIPPEC, sobre la base de OCDE - PISA 2000 - 2015

Todavía más importante fue la mejora de los aprendizajes cuando se considera la dimensión de la equidad social. Como se observa en el **gráfico 9**, América Latina logró mejorar los aprendizajes y disminuir las desigualdades entre los cuartiles de mayor y menor nivel socioeconómico en el

cualquier posterior es válida en términos oficiales para PISA. Sin embargo, en este estudio y el previo (Rivas 2015) hemos tomado los valores de las tres áreas del año 2000 como punto de partida solo para tener una referencia más amplia, advirtiendo sobre las limitaciones de esta comparación. Si bien la propia prueba PISA no realiza la comparación intertemporal de esta manera, sí lo había hecho en publicaciones previas (OCDE, 2003; OCDE, 2006; OCDE, 2009). El tomar como referencia el año 2000 para las tres áreas evaluadas permite comprender la tendencia completa de los países, sin el sesgo de un año más arbitrario para cada área evaluada, pero requiere esta aclaración metodológica: en la versión actual de PISA estos resultados no son considerados comparables, aunque sí lo fueron en publicaciones previas.

período 2000 - 2012¹¹. Al mismo tiempo, hubo regiones donde sus países mostraron una tendencia hacia el aumento de la equidad pero disminuyendo la calidad (países anglosajones) o donde bajó la equidad (Asia Pacífico).

Gráfico 9. Variación del puntaje entre PISA 2000 y 2012 (promedio lectura y matemática) y de la brecha entre resultados de los cuartiles I y IV (*), por región



(*) Para el puntaje del cuartil 1 y del 4 se considera a los resultados de los estudiantes del cuartil más bajo y más alto (respectivamente) según el índice del status económico, social y cultural de la población (ESCS, por sus siglas en inglés). Como la OCDE recalcula los índices de todos los años en cada publicación, se utilizaron para todos los años los valores informados en 2012.

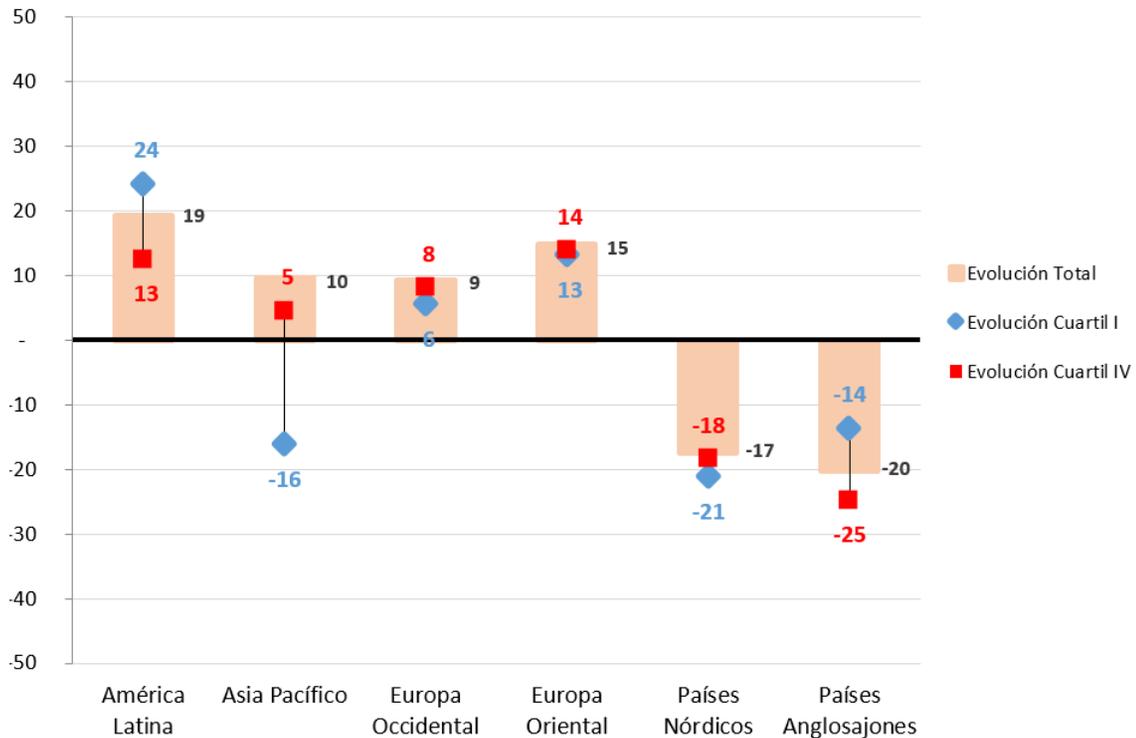
Las regiones presentadas están compuestas por los países para los cuales había datos por cuartil socioeconómico para 2000 y 2015 (con excepción de Uruguay y Colombia, que para el año 2000 se utilizaron 2003 y 2006, respectivamente): América Latina: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Alemania, Austria, Bélgica, España, Francia, Grecia, Irlanda, Italia, Liechtenstein, Luxemburgo, Países Bajos, Portugal y Suiza. Europa Oriental: Bulgaria, Hungría, Letonia, Polonia, Fed. Rusa y Rep. Checa. Países Anglosajones: Australia, Canadá, EE.UU. Nueva Zelanda y Reino Unido. Asia Pacífico: Corea y Hong-Kong. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia

Fuente: CIPPEC, sobre la base de OCDE - PISA 2000, 2003, 2006 y 2012

¹¹ El escenario que perfilan los resultados reescalados no sólo modifica los desempeños generales de los países y las tendencias en el tiempo. Es muy probable que las brechas de equidad también se vean afectadas. Como se explica en el capítulo 2, las diferencias entre ambas mediciones se vinculan con la presencia de evaluaciones con ítems no completados al final. Es probable que la metodología implementada a partir de 2015 beneficie a los estudiantes con más bajo desempeño. Lamentablemente no se dispone de información al respecto y por ello el gráfico solo alcanza hasta el año 2012.

El siguiente gráfico permite ver estas mismas dimensiones de otra forma. Allí se observa que los alumnos del cuartil de nivel socioeconómico más bajo mejoraron 24 puntos en promedio en lectura y matemática entre 2000 y 2012. Esta mejora fue superior a la de los alumnos del cuartil de más alto nivel socioeconómico de América Latina y fue superior a la de otras regiones.

Gráfico 10. Variación de puntaje PISA 2000 - 2012 (promedio lectura y matemática). Cuartil I, Cuartil IV (*) y total de evaluados, por región



(*) Para el puntaje del cuartil 1 y del 4 se considera a los resultados de los estudiantes del cuartil más bajo y más alto (respectivamente) según el índice del status económico, social y cultural de la población (ESCS, por sus siglas en inglés). Como la OCDE recalcula los índices de todos los años en cada publicación, se utilizaron para todos los años los valores informados en 2012.

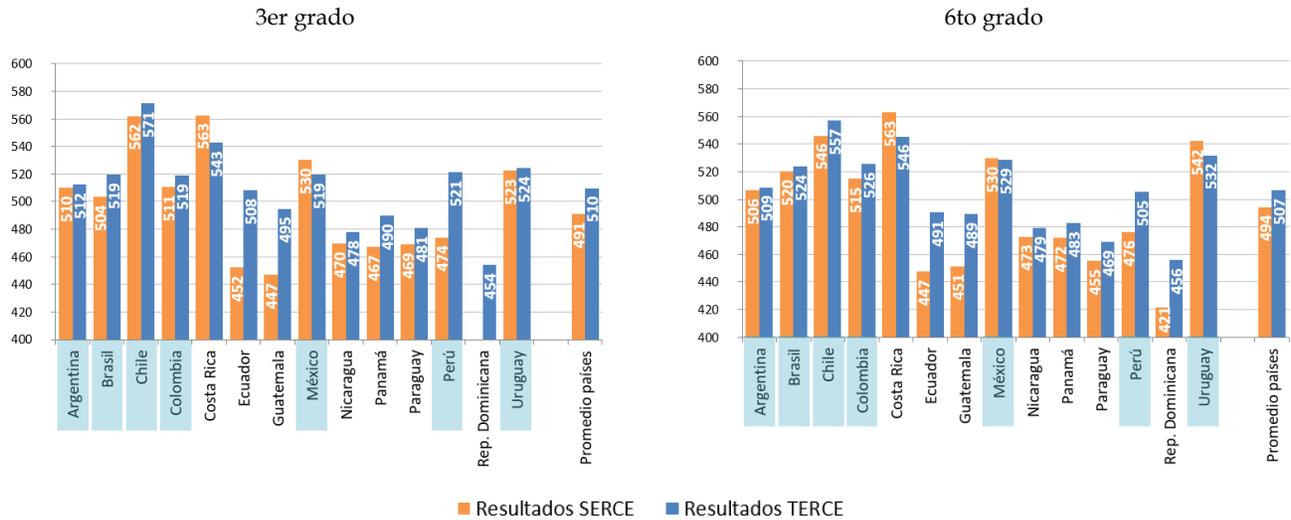
Las regiones presentadas están compuestas por los países para los cuales había datos por cuartil socioeconómico para 2000 y 2015 (con excepción de Uruguay y Colombia, que para el año 2000 se utilizaron 2003 y 2006, respectivamente): América Latina: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Europa Occidental: Alemania, Austria, Bélgica, España, Francia, Grecia, Irlanda, Italia, Liechtenstein, Luxemburgo, Países Bajos, Portugal y Suiza. Europa Oriental: Bulgaria, Hungría, Letonia, Polonia, Fed. Rusa y Rep. Checa. Países Anglosajones: Australia, Canadá, EE.UU. Nueva Zelanda y Reino Unido. Asia Pacífico: Corea y Hong-Kong. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia

Fuente: CIPPEC, sobre la base de OCDE - PISA 2000, 2003, 2006 y 2012

Esta tendencia de mejora de la calidad medida por las pruebas PISA va en la misma dirección que los resultados de las evaluaciones del LLECE-UNESCO, que mide los logros de aprendizaje en lengua, matemática y ciencias de los alumnos de tercer y sexto grado del nivel primario. En las dos pruebas comparables, SERCE de 2006 y TERCE de 2013, los 15 países participantes tuvieron en promedio una mejora destacada, especialmente en matemática en tercer grado. Como se puede

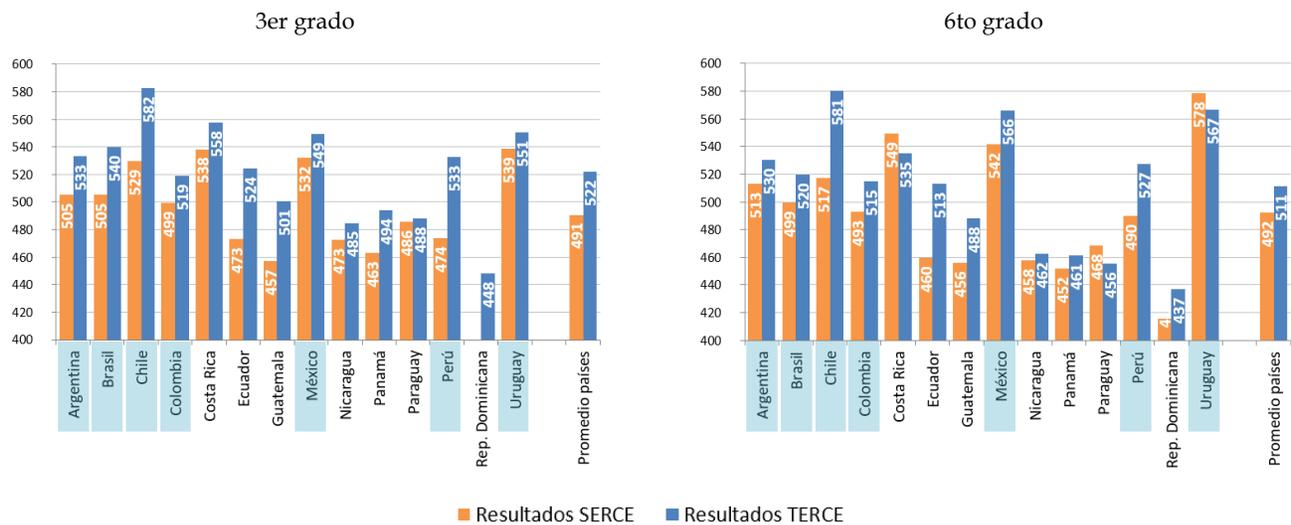
observar en el siguiente gráfico, América Latina muestra tendencia a la mejora en el desempeño de primaria entre ambas aplicaciones, en las tres áreas evaluadas y tanto para 3er grado como para 6to.

Gráfico 11. Resultados de las evaluaciones SERCE y TERCE, área de lectura, 3er y 6to grado. 2006 - 2013



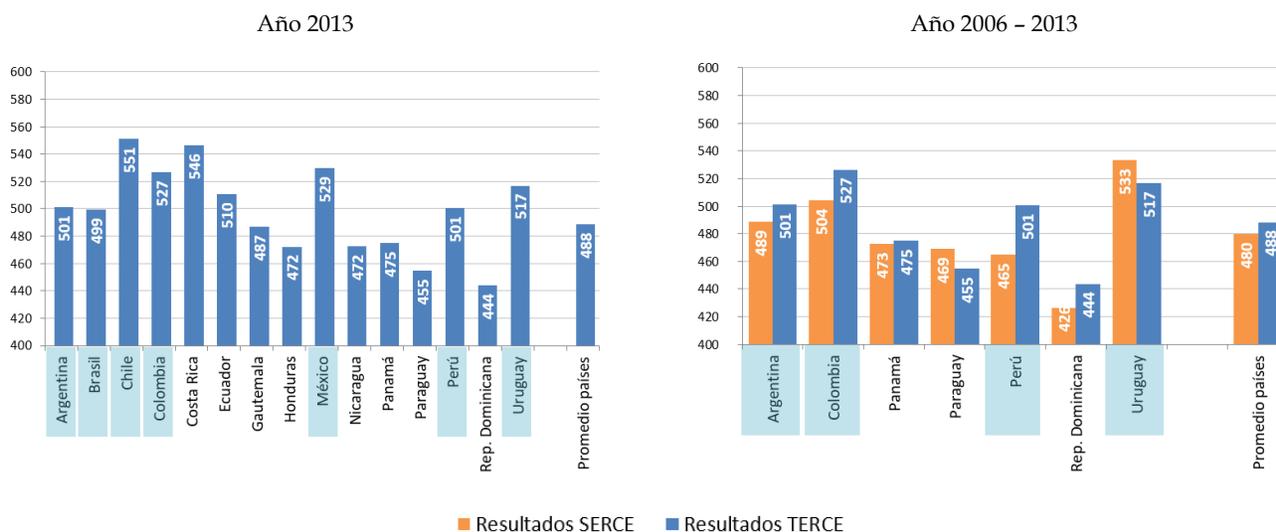
Fuente: CIPPEC sobre la base de LLECE-UNESCO (2014)

Gráfico 12. Resultados de las evaluaciones SERCE y TERCE, área de matemática, 3er y 6to grado. 2006 - 2013



Fuente: CIPPEC, sobre la base de LLECE-UNESCO (2014)

Gráfico 13. Resultados de las evaluaciones SERCE y TERCE, área de ciencias, 6to grado. 2013 y 2006 - 2013



Fuente: CIPPEC, sobre la base de LLECE-UNESCO (2014)

En el estudio previo ya citado (Rivas 2015) se analizaron en detalle estas tendencias, mostrando cómo estas mejoras en la inclusión, equidad y calidad educativa se dieron en una etapa de gran crecimiento económico, disminución de la pobreza y ampliación del financiamiento educativo que se repitió en la mayoría de los países de América Latina.

En los cuadros 8 y 9 y los gráficos 93 y 94 del Anexo I se observa que el PBI per cápita creció de \$ 11,1 mil (dólares internacionales 2011 constantes) en 2000 a \$ 14,4 mil en 2014, la pobreza disminuyó del 42,8% al 23,3% de la población (como % que vive con menos de 4 US\$ por día) y el financiamiento educativo pasó de representar el 4,06% del PBI al 5,44%. Si bien esta etapa de oro comenzó a enfriarse en los años recientes en algunos países, resulta central para comprender el contexto en el cual se analizan los resultados educativos.

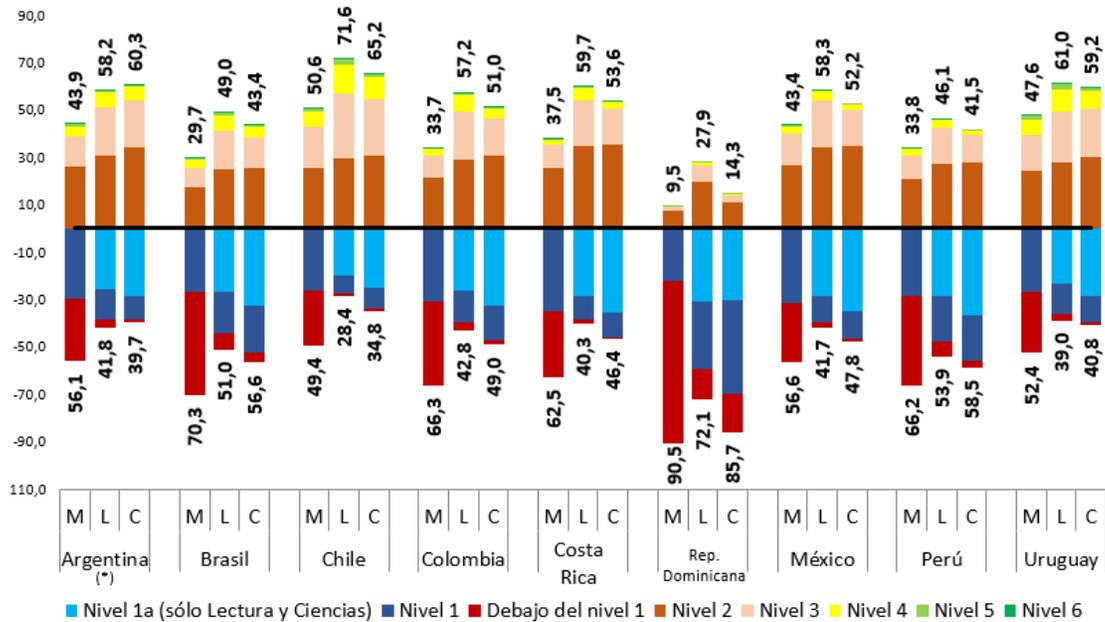
1.2. La comparación de los países

La mirada comparada de los países permite agregar una capa de profundidad al análisis de la situación educativa de la región, centrada en este primer apartado en las pruebas PISA. En este trabajo de sistematización se decidió seguir el mismo criterio que en el estudio anterior (Rivas 2015) y evitar los rankings comparados internacionales¹². No es el objetivo de este trabajo encontrar los mejores casos internacionales ni utilizar los rankings de manera inadecuada. Dado que no participan todos los países del mundo, ni una muestra representativa de ellos, el sesgo de los rankings es especialmente peligroso para comprender la posición de los países de América Latina. Creemos que

¹² Otros autores han señalado el problema que caracteriza a los rankings de países que elabora la prueba PISA. Carnoy (2015) ha criticado que no toman en cuenta los contextos sociales y económicos de los países y que simplifican su exposición al no dar cuenta de la alta magnitud del margen de error del resultado que iguala a muchos países en una misma posición relativa.

para esto resulta más apropiado presentar la situación global de la región frente a otras regiones con una contextualización de los resultados (como se ha propuesto en el apartado 1.1), y la comparación entre los países que sí parecen ser más cercanos por sus contextos tanto sociales y económicos como por sus tradiciones relativamente compartidas en aspectos educativos y culturales, tal como se propone en este apartado.

Gráfico 14. Porcentaje promedio de estudiantes en cada nivel de desempeño en Matemática (M), Lectura (L) y Ciencias (C) en PISA 2015. Países seleccionados.



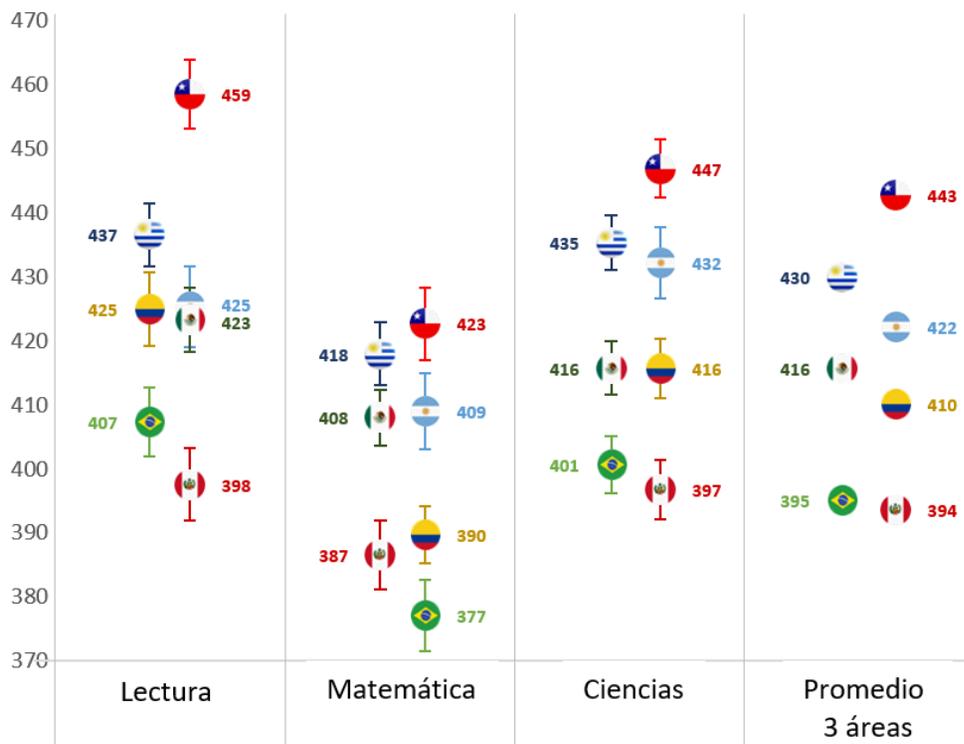
Referencias: M: matemática; L: lectura; C: ciencias.

Nota: el gráfico fue construido alineando los países en la frontera entre los niveles 2 y 1.

(*) Los resultados de Argentina 2015 son incluidos a modo ilustrativo, ya que no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: OCDE, PISA 2015

Gráfico 15. Puntaje PISA 2015 e intervalos de confianza(*), por área y promedio tres áreas. Países seleccionados (**)



(*) Se representa con barras ascendentes y descendentes, permite reconocer cuándo una diferencia es estadísticamente significativa

(**) Los resultados de Argentina fueron incluidos en el cálculo a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC, sobre la base de OCDE - PISA 2015

La edición 2015 de PISA muestra las diferencias en los resultados de aprendizaje entre los países de la región. Como se observa en el **gráfico 14**, Chile es el país que logra resultados más altos en las tres áreas con una distancia más significativa en lengua, con un 71,6% de alumnos por encima del nivel 2 (considerado como piso básico de los aprendizajes esperados). Uruguay es el país que le sigue, con resultados muy similares a Chile en matemática (ambos con cerca del 50% de alumnos por encima del nivel 2). Los resultados de Uruguay son más parejos entre las tres áreas evaluadas.

A estos dos países les siguen México y Costa Rica. En el caso de México se observa que apenas el 43,4% de los alumnos sobrepasa el nivel 2 de matemática, el 58,3% en lengua y el 52,2% en ciencias. Costa Rica está muy levemente por arriba en estas últimas dos áreas pero más abajo en matemática.

Les sigue Colombia, que logra resultados similares a México en lengua y ciencias pero más bajos en matemática, donde apenas el 33,7% de los alumnos logra alcanzar el nivel 2. Más abajo se encuentra Brasil, que obtiene puntajes apenas superiores a Perú en lengua y ciencias, pero levemente inferiores en matemática. En Brasil sólo el 29,7% alcanza el nivel 2 en matemática, el 49% en lengua y el 43,4% en ciencia. En Perú el 33,8% de los alumnos se sitúa por encima del nivel 2 en matemática, el 52,2% en ciencia y el 58,3% en lengua. República Dominicana, que participó por primera vez en PISA en 2015, alcanza resultados muy bajos.

El caso de Argentina merece un párrafo aparte dado que su participación en PISA 2015 no alcanzó los estándares metodológicos requeridos por PISA y fue excluida de la publicación central de los resultados. El caso es analizado en detalle en el Anexo II, donde se detalla que su exclusión se debió a un problema con el marco muestral entregado por Argentina a PISA. Dado que el problema del marco muestral no permite discernir si la muestra es representativa del país, los resultados de Argentina en 2015 no son comparables ni en el tiempo ni frente a otros países.

Si los datos de Argentina fuesen comparables, se podría observar que el país se situaría por debajo de Chile en todas las áreas y en una situación bastante similar a Uruguay en ciencias, pero un poco más abajo en matemática y lectura. El país estaría por encima de Colombia (excepto en lectura), de Brasil y Perú, y de México sólo en el caso de ciencias, algo que cambia la tendencia observada en las ediciones anteriores, cuando Argentina figuraba en posiciones más críticas.

Los resultados en contexto

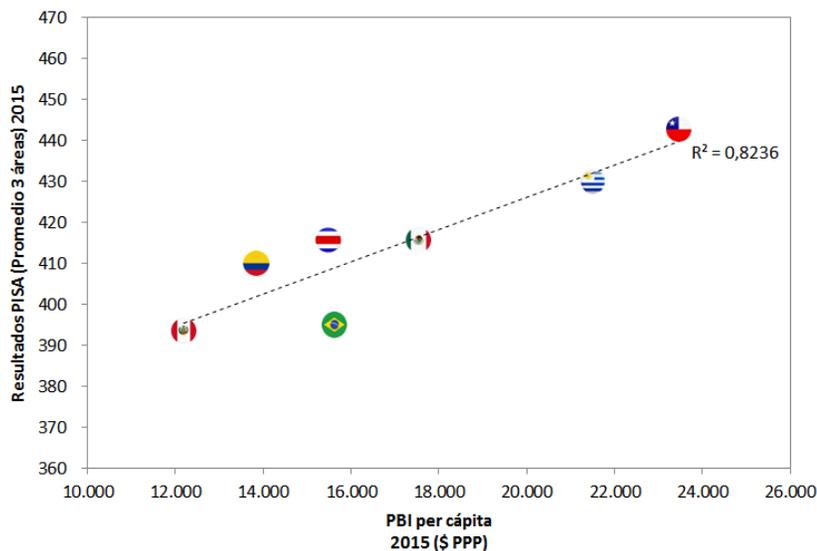
Al igual que en la comparación de regiones del mundo, es importante añadir a esta primera mirada sobre los resultados educativos en PISA, una visión más global sobre los contextos. Como marco de referencia para esta mirada, se ha elaborado una compilación de diferentes indicadores sociales y educativos, que se incluye en el Anexo I. Allí se pueden analizar datos comparados de la evolución de los contextos sociales, políticos y económicos de los países.

El **gráfico 16** muestra la relación entre los resultados en la prueba PISA 2015 (el promedio de las tres áreas evaluadas) y el PBI per cápita de los países de la región. La posición de los países en el gráfico es expresión de la relación que se presenta entre mejores resultados con mayor nivel económico¹³. En esta relación, se destaca que algunos países obtienen mejores desempeños que otros con diferente PBI, o en el caso de Colombia, Costa Rica y México, resultados similares en economías con diferentes niveles de PIB¹⁴.

¹³ No se trata aquí de hacer énfasis en la asociación estadística entre variables, ya que un correcto abordaje requiere de un análisis más desagregado e involucrando diferentes variables. El valor del r^2 se presenta a modo indicativo de la forma que asume la vinculación entre ambas variables.

¹⁴ Se podría agregar Argentina a esta lista, pero dado los problemas metodológicos ya mencionados, aquí solo figura a título ilustrativo.

Gráfico 16. PBI per cápita y resultados PISA (promedio de las 3 áreas), países seleccionados. 2015

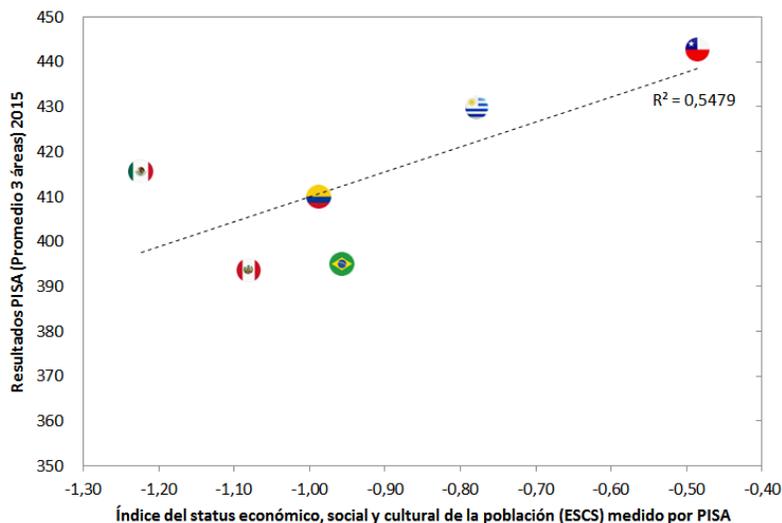


El PBI per cápita está expresado en dólares internacionales corrientes, calculados en base a la paridad del poder adquisitivo (PPA).

Fuente: CIPPEC, sobre la base de OCDE-PISA 2015 y WEO-IMF.

En forma similar, también es posible identificar que existe una vinculación entre los resultados de PISA y los niveles de pobreza de los países (medidos como valor promedio del índice ESCS de los países evaluados). Si bien la correlación es alta entre ambas variables, algunos países muestran resultados altos en relación a su nivel socioeconómico (México) o a la inversa, resultados inferiores (Brasil). Estos datos permiten poner en cuestión las miradas deterministas sobre los aprendizajes, y reforzar la necesidad de articulación de diferentes indicadores para entender mejor cada contexto nacional.

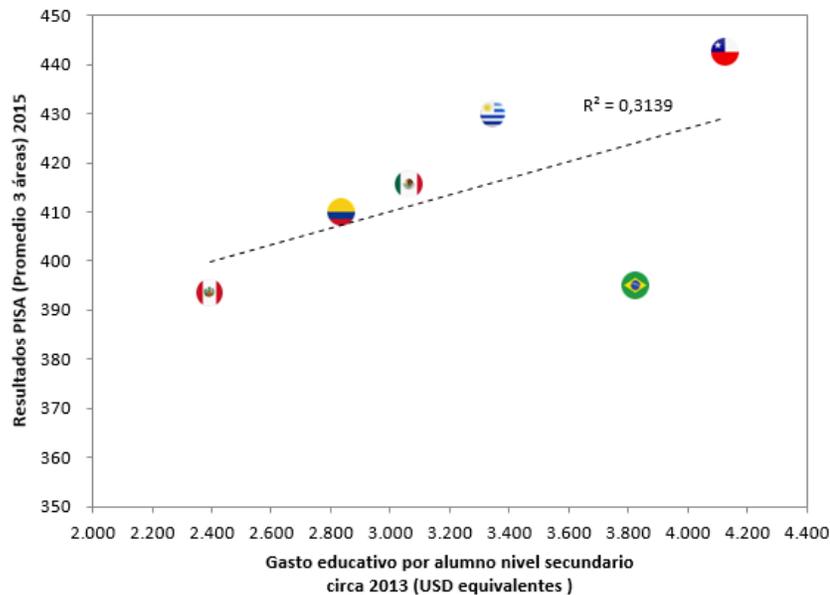
Gráfico 17. Índice del status económico, social y cultural y resultados PISA (promedio de las 3 áreas evaluadas). Países seleccionados. 2015



Los resultados de Argentina no fueron incluidos ya que la muestra no cubrió la población total objetivo, a causa de una posible omisión de escuelas del marco muestral y sus resultados no se consideran comparables.
Fuente: CIPPEC, sobre la base de OCDE-PISA 2015

En los países evaluados parece manifestarse una relación entre gasto educativo y resultados de PISA. Como puede observarse en el **gráfico 18**, el valor del índice de correlación muestra cierta relación lineal entre resultados y desempeños. Brasil es el único país que parece desprenderse de la tendencia, evidenciando altos niveles de inversión por alumno y bajos resultados.

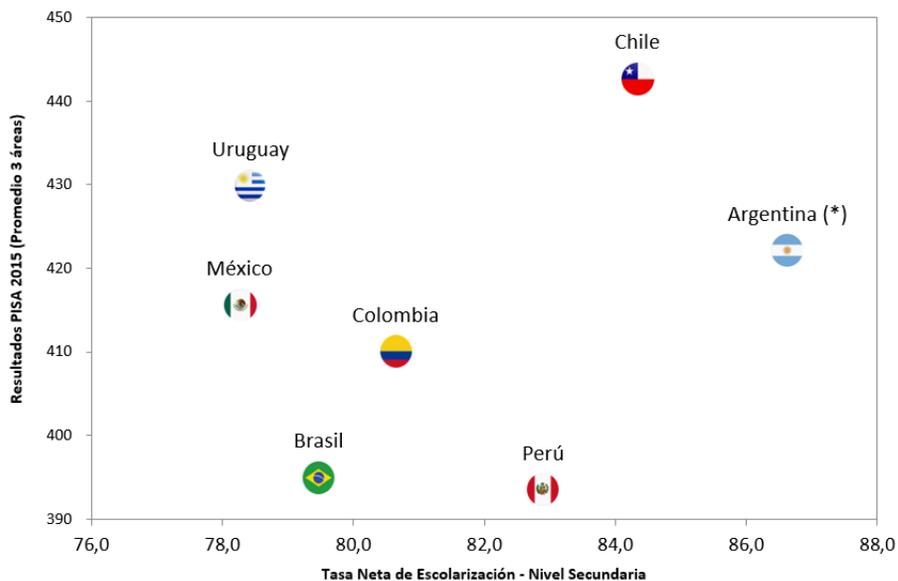
Gráfico 18. Gasto educativo por alumno de nivel secundario y resultados PISA (promedio de las 3 áreas evaluadas). Países seleccionados. Circa 2015



Nota: el gasto educativo por alumno de nivel secundario refiere al promedio del gasto público general (corriente, de capital y transferencias) en el nivel secundario, por alumno, expresado en valor nominal de dólares calculados en paridad de poder adquisitivo (PPA). Los años de referencia de los datos de gasto público por alumno son los últimos disponibles: 2014 para Perú y Uruguay, y 2013 para el resto de los países. En Brasil y Uruguay se considera únicamente el gasto por alumno de estudiantes en escuelas públicas. Los resultados de Argentina no fueron incluidos debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.
Fuente: CIPPEC, sobre la base de OCDE-PISA 2015.

Esta mirada puede ampliarse con otro dato central de los sistemas educativos: el nivel de inclusión en el nivel secundario o medio. En el **gráfico 19** se observa el cruce entre los resultados obtenidos en la prueba PISA 2015 y la tasa de escolarización del nivel secundario. Allí se puede observar que Chile se destaca por sus niveles altos de cobertura y desempeños en relación al resto de los países de la región. Perú manifiesta amplia inclusión, pero grandes desafíos en términos de aprendizajes. Uruguay, México y Brasil se sitúan en la franja inferior en los índices de cobertura, con amplias diferencias en sus resultados de PISA.

Gráfico 19. Resultado evaluaciones PISA 2015 (promedio 3 áreas) y tasa neta de escolarización en el nivel secundario. Países seleccionados. 2000-2015



(*) Los resultados de Argentina son incluidos en el gráfico a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Nota: las tasas netas de educación secundaria circa 2000 y 2015 se tomaron de SITEAL, IPE-UNESCO y OEI (Argentina, Brasil México y Uruguay), SEDLAC-Banco Mundial (Chile, Colombia), y Estadística de la Calidad Educativa - ESCALE (Perú).

Fuente: CIPPEC, sobre la base de OCDE - PISA 2015, SEDLAC, SITEAL, ESCALE (Perú).

2. Los resultados de América Latina en el tiempo: problemas y alertas sobre su comparabilidad

El siguiente paso es analizar la evolución en el tiempo. Aquí nos encontramos con los problemas metodológicos ya advertidos para estudiar la comparación de los países de América Latina en su evolución en las pruebas PISA. Dada esta situación se optó por elaborar un apartado específico explicando los cambios en la metodología de las pruebas PISA en su edición 2015 y sus implicancias para la comparación inter-temporal. En este apartado se presentan los datos de los países con las explicaciones correspondientes.

2.1. La comparabilidad de los resultados 2015

PISA introdujo en 2015 ciertos cambios en el escalamiento de los resultados con el objetivo de mejorar la precisión de las mediciones y reducir los factores de incertidumbre. Un cambio de metodología siempre genera un impacto en la comparabilidad de los resultados, que puede ser de magnitud pequeña y no afectar las tendencias, o puede ser significativo. Los cambios introducidos en 2015 son de diferente tipo, y afectaron con distinta intensidad a los países. Si bien se mantuvieron las escalas y los ítems de anclaje que garantizan la comparabilidad, fueron introducidos cuatro conjuntos diferentes de cambios que podrían incidir en la comparación¹⁵:

- Cambios en el modo de implementación: la generalización de la aplicación de PISA por computadora a casi todos los países participantes en 2015 (sólo 15 países mantuvieron la evaluación en papel) obligó a la revisión de las equivalencias entre las evaluaciones en soporte papel y las informatizadas, tanto para las comparaciones entre países como para los análisis de tendencia en el tiempo. Durante la implementación de la prueba piloto en 2014 se realizó un testeo de ítems en ambos formatos, garantizando la comparabilidad entre evaluaciones de diferente soporte¹⁶.
- Cambio en el modelo de escalamiento: al escalar los parámetros del ítem, se adoptó un modelo estadístico más flexible, de dos parámetros, distinto al utilizado en las evaluaciones anteriores, que escalaban los ítems bajo modelos de un parámetro. Este cambio permite calcular con mayor precisión la contribución de cada ítem a la estimación total del desempeño.
- Cambio en el tratamiento del modo de funcionamiento diferencial de ciertos ítems entre países. En algunos países, los ítems pueden manifestar un funcionamiento diferente al parámetro de dificultad observado en la mayoría, debido a ciertas características del idioma o del contexto cultural. En los ciclos anteriores de PISA, se utilizaban parámetros comunes y se eliminaban los ítems que mostraban variaciones significativas que podían asociarse al contexto. En 2015, la calibración permitió incorporar ítems cuyos parámetros manifestaban en ciertos contextos un funcionamiento diferente.
- Cambios en el tratamiento de los ítems incompletos al final de la evaluación: en las aplicaciones PISA anteriores a 2015, el conjunto de ítems que quedaba en blanco al final de una evaluación se consideraban como erróneos al momento de estimar el resultado del estudiante. En esta última

¹⁵ Para mayor referencia, puede consultarse el anexo A5 de OCDE (2016).

¹⁶ El “Informe sobre el estado de la educación en Uruguay 2015-2016” (INEEd, 2017) hace referencia al problema de comparabilidad de los resultados considerando el cambio en el soporte de la evaluación. Allí se remite a un estudio de Robitzsch y otros (2016) donde se analizan posibles efectos de la generalización del uso de prueba en computadora, que explicaría la caída generalizada en los desempeños en la mayoría de los países.

aplicación, se optó por considerarlos como no administrados. Este cambio metodológico tuvo un impacto decisivo en América Latina.

PISA estudió el efecto del conjunto de estas modificaciones, y estableció ciertas estimaciones de impacto en los países. Por un lado, calculó las correlaciones entre los resultados originales de cada aplicación de PISA y los obtenidos mediante esta nueva metodología. Todas son superiores a 0,993 (a excepción del área de lectura en 2006, donde es de 0,985), lo que indica una alta correspondencia entre ambas formas de estimar los desempeños.

Por otro lado, calculó el error de enlace entre todas las áreas y años respecto del 2015, que se estima en torno a los 5 puntos. Se concluyó que estas modificaciones, en su conjunto, inciden levemente en la comparabilidad entre países, y se asumió un margen de incerteza en la comparación entre resultados de aproximadamente 10 puntos.

Sin embargo, al analizar en detalle el impacto de estos cambios en los países de América Latina, la incidencia parece ser mayor. Ya algunos informes nacionales de PISA, como el caso de México, establecen algunas advertencias sobre posibles limitaciones en la comparación asociadas a estas modificaciones (INEE, 2016). Uruguay necesitó publicar una nota metodológica específica para explicar los problemas en la comparabilidad de los resultados del país (ANEP, 2016). Es necesario detenerse en este análisis, ya que tiene particulares implicancias para América Latina, y amerita un tratamiento específico en este documento.

En este punto, es importante detenerse a aclarar que la propiedad de comparación entre relevamientos a lo largo del tiempo, e incluso entre países, nunca es absoluta. A toda estimación estadística siempre hay asociado un error. Siempre existe un marco de incertidumbre o incerteza, debido a que nunca es posible replicar en forma precisa los medios y condiciones mediante los cuales se obtiene información en dos aplicaciones diferentes. Por esta razón, se establecen ciertos criterios rigurosos de procedimientos para la construcción e implementación de los instrumentos que se orientan a disminuir esta incertidumbre.

Hay ciertos errores que ejercen incidencia en las comparaciones, y que son calculables. PISA considera algunos de ellos y los analiza. Un caso es el de la comparación en el tiempo de cada área: no todas son comparables con el mismo nivel de precisión. En cada aplicación PISA se hace énfasis en un área, que va rotando de ciclo en ciclo. Este énfasis está dado porque se incluyen más ítems específicos del área en la evaluación, en detrimento de las otras dos¹⁷. De todas formas, es un tipo de error que es mensurable, y PISA incluye en las publicaciones la estimación de los márgenes de error en las comparaciones en el tiempo.

Otro error en las comparaciones, también calculable, ocurre cuando por el paso del tiempo la población objetivo se ve afectada por cambios sociodemográficos. En los análisis sobre las tendencias a lo largo del tiempo, PISA suele calcular cuál sería el comportamiento de los resultados si se

¹⁷ En la primera aplicación de PISA del 2000 se hizo énfasis en lectura, en la segunda del 2003 en matemática, en la tercera del 2006 en ciencias, y en las subsiguientes se mantuvo este esquema rotativo, llegando a la aplicación 2015 con énfasis en ciencias. Debido a esto, la comparabilidad en el tiempo se inicia para cada área en el primer año donde fue considerada como prioritaria (lectura desde el 2000, matemática desde el 2003, ciencias desde el 2006). Asimismo, la comparación entre años donde se hizo énfasis en una misma área (por ejemplo, matemática entre 2003 y 2012) es más precisa que cuando se compara entre otros ciclos, porque es mayor la cantidad de ítems de anclaje involucrados en la comparación. Cabe señalar que en PISA 2015 se introdujeron modificaciones en la distribución de ítems entre las tres áreas, para que las comparaciones en el tiempo sean más robustas para aquellas áreas “menores” de la aplicación.

mantuviera la misma proporción de población por sexo, edad y condición migratoria, y establece si existen cambios significativos. Estos ejercicios se realizan a modo ilustrativo, pero no se toman en cuenta para la construcción de los resultados de cada año de aplicación. En general, la incidencia de estos factores suele ser pequeña, situándose por debajo de los 5 puntos¹⁸ en la escala PISA (que se normaliza en una media de 500 y desvío de 100). Estas diferencias no son incorporadas al cálculo del error de comparación.

En otras ocasiones, los errores se manifiestan en alguna etapa del proceso de construcción del dato, por alguna falla del proceso o por alguna particularidad del sistema educativo no contemplado en la muestra. Estos son más problemáticos, porque no se puede calcular el impacto de este tipo de error. La detección de estos errores de implementación de una aplicación llevan a determinar que los resultados tienen (o podrían tener) sesgos, y se establece que no son comparables, afectando la serie de tiempo.

Algunas de estas dificultades son reconocidas en los informes de PISA. Es posible mencionar algunos ejemplos tomando como referencia los datos publicados en el Volumen V del informe PISA 2009 (OCDE, 2010), cuya temática central es el análisis de las tendencias en el tiempo de los resultados. Allí se incluyen algunas menciones a casos de países donde se ha visto afectada la comparabilidad.

El problema de comparabilidad que se aborda en este apartado es de naturaleza diferente a los anteriormente descritos. Lo significativo del caso es que no resulta de errores o inconsistencias de la aplicación de la evaluación en un país, sino que su origen está en los cambios en la metodología de construcción de resultados que la propia evaluación implementó.

El informe PISA 2015 permite el acceso a los resultados 2012 - 2015 recalculados en base a estos cambios. A partir de pedidos formales a la OCDE, en este estudio se accedió a los resultados reescalados comparables para las evaluaciones de los años 2006 y 2009. En el gráfico que sigue, se presentan estos resultados reescalados junto con los "originales" de cada aplicación, a los fines de dimensionar en qué medida se ve afectada la comparación en el tiempo en las distintas regiones.

Se muestran los resultados de los ciclos 2006 - 2009 - 2012 - 2015 bajo dos metodologías: (a) la comparación directa entre los resultados originales de cada año de aplicación, y (b) el reescalamiento de los resultados de años anteriores implementando los cambios introducidos en 2015.

¹⁸ Para mayor referencia, puede consultarse el gráfico I.2.23 de OCDE (2016) PISA 2015 Results (Volume I) Excellence and Equity in Education.

Gráfico 20. Resultados PISA 2006 - 2015, puntajes originales y reescalados (promedio tres áreas), por región¹⁹



Nota: en América Latina se excluye Perú, ya que no tiene aplicación PISA en 2006, y a Argentina, ya que los resultados 2015 no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

El rango de la escala de los gráficos se ajusta a los resultados de cada región, pero se mantienen amplitudes equivalentes entre los gráficos para que las brechas sean comparables

Las regiones se construyeron con el promedio simple de países participantes en todos los años de PISA, considerando la siguiente asignación: América Latina: Brasil, Chile, Colombia, México y Uruguay. Asia Pacífico: Taipéi (China), Hong Kong (China), Macao (China), Japón y Corea. Europa Occidental: Austria, Bélgica, Francia, Alemania, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Portugal, España y Suiza. Europa Oriental: Bulgaria, Croacia, Rep. Checa, Estonia, Hungría, Letonia, Lituania, Montenegro, Polonia, Rumania, Fed. Rusa, Rep. de Eslovaquia y Eslovenia. Países Anglosajones: Australia, Canadá, Nueva Zelanda, Reino Unido y Estados Unidos. Países Nórdicos: Dinamarca, Finlandia, Islandia, Noruega y Suecia.

Fuente: CIPPEC sobre la base de PISA 2015

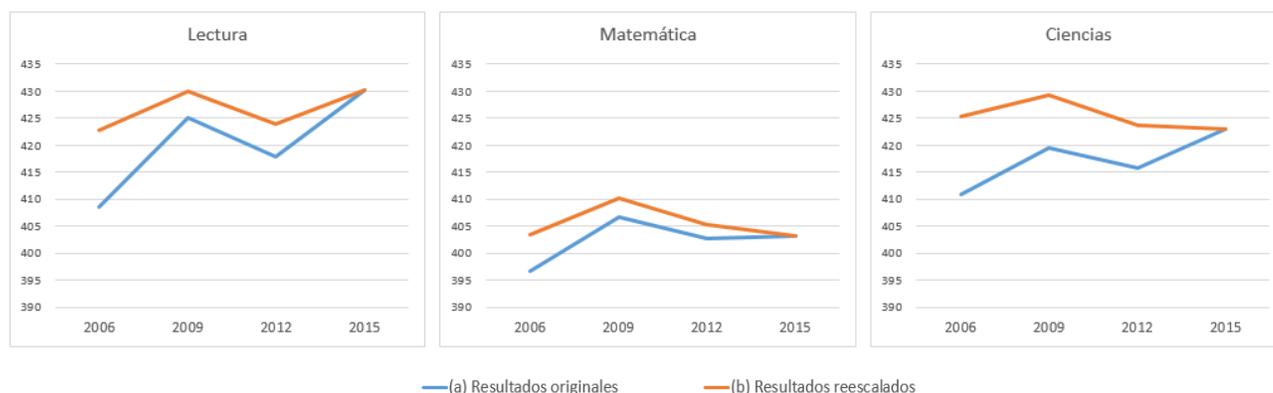
Se observa con claridad que estos cambios introducidos en el escalamiento de los puntajes de PISA 2015 afectan fuertemente la comparabilidad de los resultados de América Latina, mientras que en el resto de las regiones las variaciones son casi imperceptibles.

19 Los países considerados para obtener los resultados de América Latina que se presentan en los gráficos de este apartado difieren de los incluidos en otros capítulos del informe. Se decidió aquí trabajar exclusivamente con los países que poseen datos completos y válidos en toda la serie 2006-2015 para construir comparaciones metodológicamente más rigurosas, y por esta razón se excluye también a Perú del cálculo, que no participó en PISA 2006. En los otros gráficos del documento se incluye con una estimación lineal basada en la evolución 2000 - 2009.

Obsérvese el gráfico superior izquierdo: tomando en cuenta los resultados publicados en cada aplicación, es reconocible una mejora en América Latina, de un promedio de 405 puntos en 2006 a 417 en 2015. En el período analizado, es una de las dos regiones que han mejorado sus resultados entre 2012 y 2015. A su vez, también es la que más ha crecido si se considera el período completo 2006 - 2015. Ahora bien, si se observa el resultado recalculado con la metodología de 2015 (resultados reescalados), el perfil de la región es muy distinto.

Para el año 2006, estos resultados son superiores a los “originales” en 12 puntos (teniendo en cuenta el promedio de las 3 áreas). Como los resultados de las aplicaciones 2006 - 2012 son mayores al utilizar la metodología de 2015, las tendencias de la región cambian de sentido: los resultados de PISA 2015 ya no son los más altos del período, sino muy similares a los años anteriores (con excepción de 2009). En promedio, los resultados de las tres áreas se mantienen sin variaciones en el período, a excepción de una mejora circunscripta a la aplicación 2009. Los países no han mejorado su rendimiento en PISA en los últimos nueve años. Esta diferencia puede observarse en todas las áreas evaluadas, con particular énfasis en ciencias, como se observa en el siguiente gráfico.

Gráfico 21. Resultados PISA 2006 - 2015, puntajes originales y reescalados, por área. Promedio de países de América Latina



Nota: en América Latina se excluye a Perú, ya que no tiene aplicación PISA en 2006, y a Argentina, ya que los resultados 2015 no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC sobre la base de PISA 2015

En el promedio de países considerados, puede identificarse cómo cambia el diagnóstico de cada área según se consideren los resultados “originales” (es decir, los publicados en cada aplicación), o los reescalados con la metodología utilizada en PISA 2015.

Cada caso amerita abordar el análisis de sus particularidades. En el área de ciencias se observan las brechas más marcadas entre la metodología en los años iniciales del período analizado, con una reducción marcada en las siguientes aplicaciones. En el caso de matemática, si bien las diferencias son menores, se mantienen estables. Se tabulan a continuación las observaciones que se desprenden del análisis de cada una de las series de resultados, a los fines de comparar las diferencias de diagnóstico:

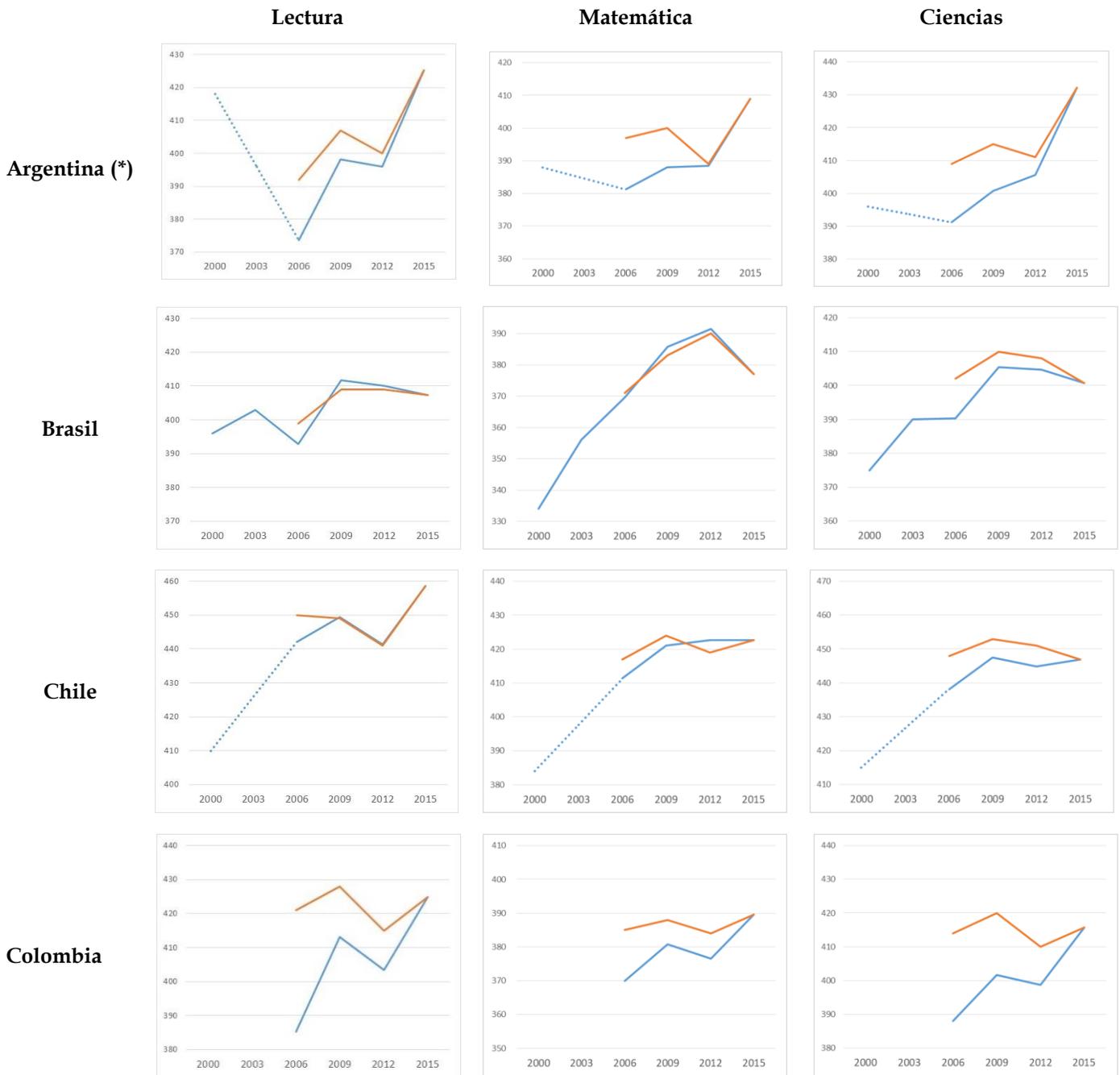
Cuadro 1. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), por área. América Latina (países seleccionados)

	<u>Resultados originales</u>	<u>Resultados reescalados</u>
Lectura	<ul style="list-style-type: none"> • La región mejoró sustantivamente entre 2006 y 2015, con un incremento promedio de 21 puntos. • Esta mejora se observa en los ciclos 2006 – 2009 y 2012 – 2015. • En el 2015 se alcanzan los resultados más altos de todo el período. 	<ul style="list-style-type: none"> • La región mejoró levemente entre 2006 y 2015, con un incremento promedio de 8 puntos. • Esta mejora se observa en los ciclos 2006 – 2009 y 2012 – 2015. • En el 2015 se alcanzan resultados similares al 2009.
Matemática	<ul style="list-style-type: none"> • Se observa una mejora 2006 – 2009, y luego un leve empeoramiento de los resultados. • En el período completo, los resultados de matemática subieron 6 puntos. 	<ul style="list-style-type: none"> • Se observa una mejora 2006 – 2009, y luego un leve empeoramiento de los resultados de 7 puntos de magnitud. • En el período completo, los resultados de matemática no se modificaron.
Ciencias	<ul style="list-style-type: none"> • La región mejoró entre 2006 y 2015, con un incremento promedio de 12 puntos. • Esta mejora se observa en los ciclos 2006 – 2009 y 2012 – 2015. • En el 2015 se alcanzan los resultados más altos de todo el período. 	<ul style="list-style-type: none"> • La región no mejoró entre 2006 y 2015, con una reducción promedio de 3 puntos. • Desde los ciclos 2009 en adelante los resultados no mejoran. • En el 2015 se alcanzan los resultados más bajos de todo el período, similares a 2012.

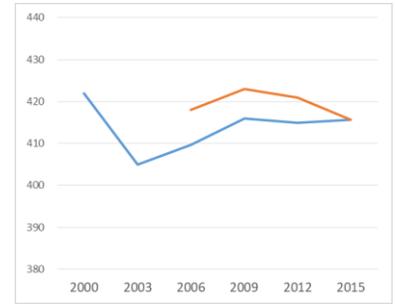
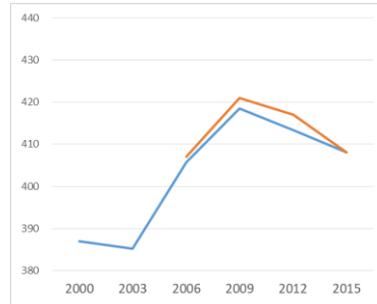
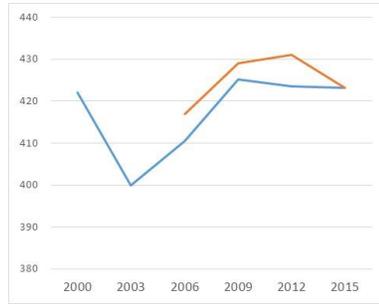
Estas variaciones no son iguales entre países y áreas. A continuación se presentan los resultados PISA de los 7 países de América Latina analizados en este estudio, para cada una de las áreas evaluadas, comparando los resultados originales y los reescalados. A los fines de analizar estas diferencias en el contexto global de las aplicaciones de PISA, se incluyen los resultados originales 2000 – 2015 de las tres áreas, y se incorpora la serie 2006 – 2015 de los resultados reescalados. Cabe recordar que en el periodo para el cual se presentan los resultados reescalados sólo el área de ciencias fue el principal dominio evaluado en dos años (2006 y 2015), mientras que lectura lo fue en 2009 y matemática en 2012²⁰.

²⁰ Como fue explicado anteriormente, en PISA 2015 fueron incorporadas modificaciones en la cantidad de ítems aplicados por área, para mejorar la precisión de las comparaciones aún en las áreas que no se consideraron como principal dominio a evaluar. Por lo tanto, si bien se considera todo el periodo para el análisis de tendencias, las comparaciones son más robustas en los periodos 2006 – 2015 para ciencias, 2009 – 2015 para lectura y 2012 – 2015 para matemática.

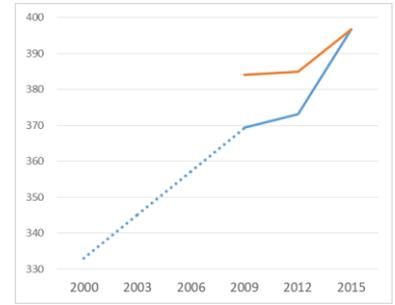
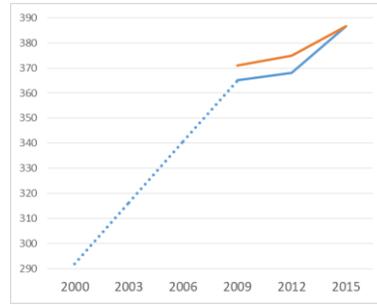
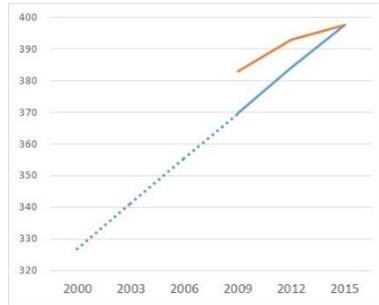
Gráfico 22. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), por área. Países de América Latina



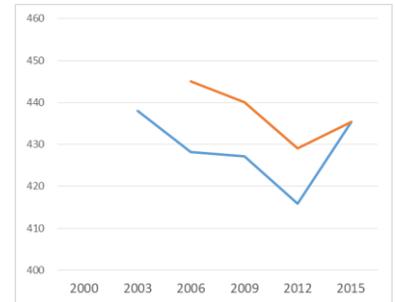
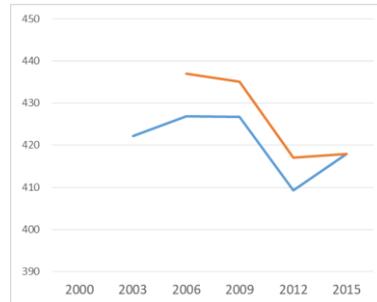
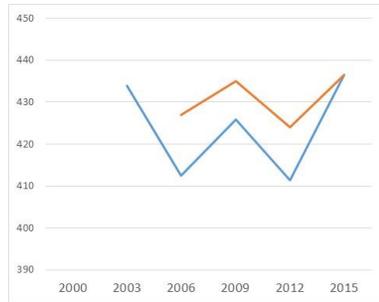
México



Perú (i)



Uruguay



— (a) Resultados originales — (b) Resultados reescalados

Nota: El rango de la escala de los gráficos se ajusta a los resultados de cada país, pero se mantienen amplitudes equivalentes entre los gráficos para que las brechas sean comparables, excepto en el caso de Perú (i), donde fue necesario ampliar el rango del gráfico para representar el incremento de los puntajes en el periodo analizado.

(*) Los datos de Argentina fueron incluidos para analizar el efecto del cambio de metodología en PISA 2006 - 2012, se incluyen los resultados 2015 aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC sobre la base de PISA 2015

Como primer comentario general de análisis, se observa que los resultados reescalados se sitúan, en prácticamente todos los casos, en valores similares o superiores a los resultados originales. Esta característica es consistente con lo observado en los promedios regionales, y tiene importantes implicancias: en los casos donde se evidencian diferencias entre ambos cálculos, se trata de países que en años anteriores han tenido resultados más altos en PISA según la metodología implementada en 2015. Por lo tanto, los desempeños han mejorado menos de lo que se estimaría al analizar los resultados publicados originalmente en cada aplicación.

En el área de **lectura**, se observa que las mayores diferencias entre la forma de calcular resultados se concentran principalmente en Colombia, con amplios márgenes entre ambos resultados en los años comparados. En el año 2006 las brechas son muy significativas, los resultados de lectura reescalados muestran 36 puntos de diferencia respecto de los resultados originales. También en Argentina y Uruguay se observan diferencias importantes, cercanas a los 15 puntos en PISA 2006, y a los 10 puntos en PISA 2009. Si bien los resultados de Argentina no muestran diferencias en 2012, en Uruguay permanecen brechas amplias en torno a los 13 puntos. En Perú, que no participó de PISA 2006, se observan diferencias en torno a los 10 puntos en 2009 y 2012. En los casos de Brasil, Chile y México, las diferencias son menores.

En el caso de **matemática**, las comparaciones entre ambos resultados permiten hacer observaciones similares, aunque con diferencias menores a las identificadas en lectura. En el 2006 Argentina y Colombia muestran las mayores diferencias, en torno a los 15 puntos entre ambos valores. En Uruguay las diferencias en este año son de 10 puntos, y en el resto de los países son menores. Las diferencias se sostienen en 2009 y 2012 para Colombia y Uruguay, y se incorpora Perú también. En Argentina para el año 2012 ya no se visualizan brechas entre ambos cálculos. Brasil, Chile y México muestran resultados similares en todos los años.

Por último, se destaca el caso de **ciencias** porque las diferencias entre los modos de calcular los resultados arrojan diferentes escenarios en casi todos los países. En el año 2006, a excepción de México, en todos los países las diferencias son superiores a los 10 puntos, cercanas a los 20 en Argentina y Uruguay, y superiores a 25 en Colombia. Cabe recordar que la comparación 2006 - 2015 en ciencias involucra dos aplicaciones de PISA donde esta área fue el principal dominio evaluado. Colombia, Perú y Uruguay siguen mostrando diferencias amplias en 2009 y 2012 (entre 12 y 18 puntos), Argentina solamente en 2009, y en Chile, México y Brasil a partir del 2009 las diferencias entre resultados se reducen a valores bajos, de entre 3 y 7 puntos.

La observación de este conjunto de gráficos ha permitido reconocer que las diferencias entre los métodos de cálculo de resultados arrojan escenarios diferentes, que varían entre países y áreas con diferente intensidad. ¿Qué implicancias tiene este cambio de metodología para interpretar el desempeño de los países en los últimos diez años de PISA?

Para algunos países y áreas, los resultados reescalados con la metodología de PISA 2015 no difieren significativamente de aquellos que fueron publicados en cada aplicación. En estos casos, se podría afirmar que los diagnósticos no varían sensiblemente.

Sin embargo, hay un conjunto importante de resultados que se modifican en forma muy significativa. En algunos países, con la metodología anterior se observaba una mejora sustantiva en los desempeños. Tal es el caso, por ejemplo, de incrementos que son señalados como estadísticamente significativos (Colombia en ciencias entre 2009 y 2015, Argentina en lectura entre 2006 y 2012, Uruguay en matemática entre 2012 y 2015, Perú en matemática entre 2012 y 2015). Todas estas mejoras desaparecen (se reducen a pequeñas diferencias que no son estadísticamente significativas) al recalcular los resultados con la metodología del 2015.

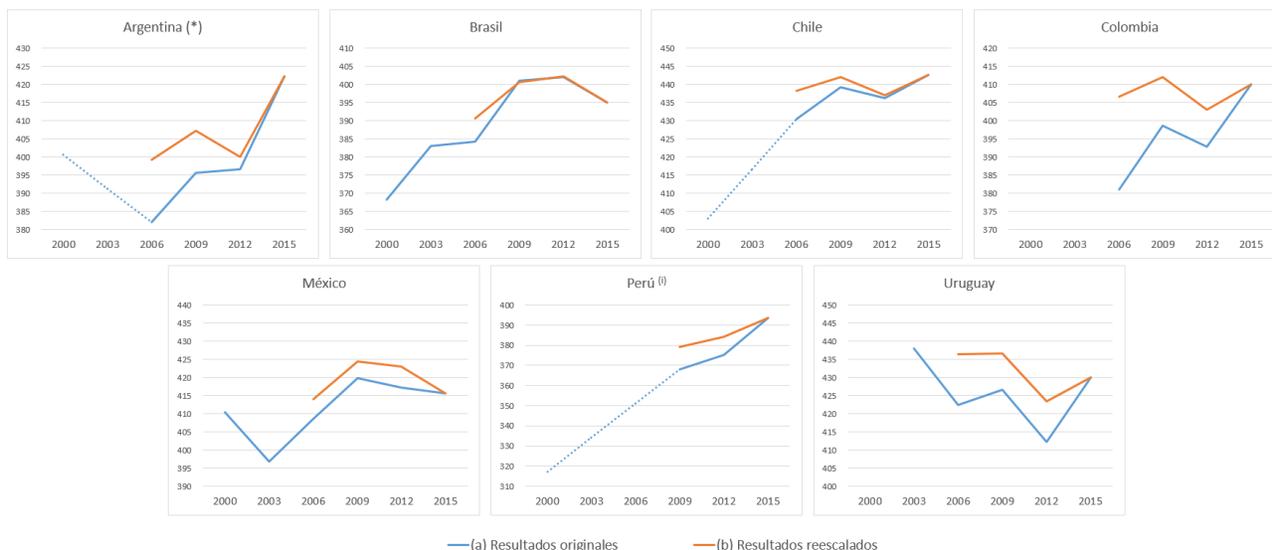
En otros casos, lo que parecía un resultado sin modificaciones, se presenta ahora como una caída en el desempeño (como en el caso de Argentina en matemática 2009 - 2012, que con los resultados reescalados muestra una disminución estadísticamente significativa), o se manifiesta una mejora que mantiene significancia estadística pero es de menor magnitud a la que arroja la comparación directa (Perú en lectura 2009 - 2015).

Un problema adicional es la falta de información sobre los resultados previos al 2006: no es posible saber cuál es el resultado con la metodología 2015 (es decir, comparable) en los primeros años de PISA (para los cuales no se han publicado los datos reescalados), periodo en el que varios de los países analizados muestran mejoras sustantivas, que han sido resaltadas como señal de los avances significativos en los países de la región en materia de aprendizajes (Rivas, 2015). Es muy posible que este cambio de metodología afecte a los resultados en estos años, y en algunos países las tendencias sean diferentes. Este problema no se agota en su aspecto metodológico. El cambio introduce interrogantes a la capacidad de PISA de cumplir de su objetivo principal: poder realizar apreciaciones sobre calidad, equidad y eficiencia de los sistemas escolares de los países participantes.

Sin embargo, es importante resaltar que el cambio de metodología permitirá un monitoreo de los aprendizajes de mayor calidad en aplicaciones futuras de PISA. Estos cambios contribuyen a la construcción de una medición más precisa de los desempeños, que permite comparaciones más robustas hacia adelante.

Una mirada sintética sobre los resultados de cada país permite identificar con mayor precisión cuáles se ven más afectados con el cambio metodológico. Se presenta a continuación un resumen de los resultados PISA de los países analizados, promediando las tres áreas evaluadas, manteniendo el mismo criterio de representación.

Gráfico 23. Resultados PISA en puntajes originales (2000-2015), y en puntajes reescalados (2006-2015), promedio tres áreas. Países de América Latina



Nota: El rango de la escala de los gráficos se ajusta a los resultados de cada país, pero se mantienen amplitudes equivalentes entre los gráficos para que las brechas sean comparables, excepto en el caso de Perú (i), donde fue necesario ampliar el rango del gráfico para representar el incremento de los puntajes en el periodo analizado.

(*) Los datos de Argentina fueron incluidos para analizar el efecto del cambio de metodología en PISA 2006 - 2012, se incluyen los resultados 2015 aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC sobre la base de PISA 2015.

Brasil es el país menos afectado por estos cambios: en promedio, los resultados de las tres áreas muestran valores similares en el periodo 2006 - 2015. En ambos casos, los resultados de PISA presentan una tendencia a la mejora entre los años 2006 y 2009 (diferencia estadísticamente significativa), y luego un retroceso en 2015. Las diferencias entre ambos escalamientos se concentran en el 2006, con una diferencia promedio de 7 puntos. Por esta razón, si se consideran los resultados originales, Brasil muestra en el período 2006 - 2015 una mejora de 11 puntos (promedio tres áreas), pero si se consideran los resultados reescalados, esta diferencia se reduce a 4, un escenario de estabilidad con diferencias que no son estadísticamente significativas en ninguna de las tres áreas.

En **México** también puede verse una tendencia similar entre ambas formas de construir los resultados. Ya sea con los resultados publicados o con los reescalados, puede reconocerse una mejora en el período 2006 - 2009 (con diferencias estadísticamente significativas en lectura y matemática), y un deterioro posterior de los resultados. La diferencia entre ambas mediciones está dada principalmente por la tendencia global del período: tomando los resultados originales de cada operativo, la mejora promedio en cada área es de 7 puntos entre 2006 y 2015. Con los resultados reescalados, esta mejora se reduce a dos puntos.

Chile es un caso de incidencia media en el cambio de la metodología: las diferencias entre ambas formas de calcular los resultados son mayores en PISA 2006, y menores en 2009. Si se consideran los resultados originales, Chile obtuvo en el 2006 los resultados más bajos del período, y evidencian una mejora hacia 2009, de 8 puntos. En cambio, con los resultados reescalados, el incremento 2006-2009 sería menos intenso (aunque con ambas mediciones con diferencias que no son

estadísticamente significativas) y los resultados 2006 serían muy similares al 2012. En síntesis, con los resultados originales, Chile mostraría una tendencia de mejora en el período que se podría poner en duda al mirar los resultados reescalados: con la metodología de PISA 2015, los resultados del último año son similares al 2009, y los del 2012 muy cercanos a los del 2006.

En **Colombia**, el escenario es muy diferente: los resultados publicados de las aplicaciones PISA 2006 - 2015 muestran una sustantiva mejora, de 381 a 410 puntos en el período (en promedio de las tres áreas, con diferencias estadísticamente significativas en todas). Ahora bien, si se analizan los resultados reescalados, el país muestra una tendencia estable en el periodo, con ciertas oscilaciones que no son significativas y no marcan una tendencia clara. Se podría construir con estos datos un escenario bien diferente: un país que no manifiesta haber avanzado en los aprendizajes medidos por PISA, manteniendo niveles similares en los últimos diez años.

En **Uruguay**, el escenario también se modifica sustantivamente. Con los resultados originales publicados, el país muestra cierta estabilidad, con una caída 2009 - 2012 y una sustantiva mejora en la última aplicación. Ahora bien, con los resultados calculados con la metodología de PISA 2015, el escenario que se perfila es inverso, con un empeoramiento de los resultados en los últimos años. Los resultados promedio del país en 2006 y 2009 (436 y 437 respectivamente) son superiores a los del 2012 (423 puntos) y al 2015 (430 puntos). La mejora estadísticamente significativa en lectura entre 2012 y 2015 se diluye, y el escenario de estabilidad en matemática se reemplaza por una disminución significativa de los desempeños.

En **Perú** la tendencia es creciente en ambos escenarios, dando cuenta de una mejora efectiva de los aprendizajes. Sin embargo, el incremento de los resultados es mucho menor con los resultados reescalados: el crecimiento 2009 - 2015 se reduce de 26 puntos a 14 puntos entre escenarios (promedio tres áreas). La utilización de la nueva metodología implementada en el 2015 implicaría afirmar que en el país la mejora identificada en los desempeños de los estudiantes ha sido mucho menor a la que permiten caracterizar los resultados publicados en una comparación directa. En matemática, el incremento 2012 - 2015 se vuelve no significativo en términos estadísticos, perfilando un escenario de estabilidad.

Por último, en el caso de **Argentina** se observa también una importante incidencia de este cambio metodológico, especialmente en los años 2006 y 2009. Como los resultados 2015 no son comparables, el análisis se enfoca en el periodo 2006 - 2012. Si se consideran los datos originales de cada implementación de PISA, Argentina mostraba, para las tres áreas, una mejora en los resultados, principalmente explicada por un incremento promedio de 14 puntos hacia el año 2009. Con los resultados reescalados con la metodología de PISA 2015, esta mejora 2006 - 2009 es de una magnitud menor, y acompañada por un deterioro posterior de los resultados, donde los puntajes obtenidos en 2012 son muy similares al 2006. De esta forma, con estos resultados Argentina pasa de perfilar una mejora a una estabilidad en su desempeño en PISA. Este cambio se intensifica en lectura, donde la diferencias estadísticamente significativas entre 2006 y 2012 se diluyen. Esto se explica porque los resultados de las pruebas anteriores a 2015 no eran tan bajos como originalmente se había publicado con la metodología anterior de PISA.

En América Latina, como fue puesto en evidencia, el cambio de metodología implementado en 2015 introduce profundos interrogantes sobre la comparabilidad de PISA en el tiempo, sembrando un manto de duda en todas las conclusiones que se han elaborado sobre los resultados 2006 - 2015, especialmente en algunos países como Colombia, Argentina, Perú y Uruguay.

Ningún análisis en el tiempo que se realice sobre los resultados de PISA puede desconocer este problema. Aun así, el informe PISA 2015 presenta información incompleta para entender el problema

de la comparabilidad de los resultados, dado que fue confinada a un anexo y no tiene ninguna otra referencia en el resto del documento.

Resulta llamativo que estas diferencias tan importantes, producidas por el cambio de metodología, no hayan sido analizadas en detalle en los informes centrales, evitando las conclusiones que se exponen tomando los datos originales sin las correcciones metodológicas. En la práctica, PISA analiza los datos de los resultados en el informe 2015 como si fuesen comparables, resaltando los desempeños de los países, y el abordaje de estos problemas de comparabilidad queda relegado a un anexo metodológico²¹.

Para que la evaluación PISA pueda brindar información que permita orientar a los países de la región en lo que respecta a la calidad, equidad y eficiencia de sus sistemas escolares (tal como se postula explícitamente en los objetivos de PISA) resulta indispensable abordar en profundidad este problema, y establecer criterios claros, válidos y confiables para abordar el análisis de tendencias en el período 2006 - 2015 en los países de la región.

2.2. ¿Por qué el cambio de metodología en PISA 2015 afecta tanto a América Latina?

Tal como fue mencionado previamente, en PISA 2015 se introducen algunos cambios en los criterios de escalamiento de los resultados, que muestran un mayor impacto en América Latina por sobre el resto de las regiones. La hipótesis que aquí se explora es que la principal razón por la que se produce este impacto diferencial en América Latina es la aplicación de un cambio en particular: el tratamiento de los ítems incompletos al final de la evaluación.

Esta modificación refiere a la forma en que se considera, para la estimación de los resultados, un tipo particular de no respuesta. En las aplicaciones PISA anteriores a 2015, los ítems que quedaban en blanco al final de una evaluación se consideraban como erróneos para estimar el resultado del estudiante²². En esta última aplicación, se optó por considerarlos como no administrados. En parte, el motivo de este cambio se vincula con la aplicación de la prueba en computadora: los ítems se presentan secuenciados, y los estudiantes nunca llegan a ver las preguntas que no abordaron al final del test ni pueden saber cuántas les quedan como para planificar y organizar su esfuerzo (OCDE 2016g).

²¹ Al respecto, uno de los casos más notables es el de Colombia. A lo largo del informe, PISA resalta en más de una ocasión los logros del país en la mejora de los desempeños en Ciencia entre 2006 y 2015, como se menciona por ejemplo al inicio del informe: *“La última vez que la ciencia fue el foco de PISA fue en 2006. (...) es decepcionante que para la mayoría de los países con datos comparables el rendimiento de la ciencia en PISA se mantuvo prácticamente sin cambios desde 2006. De hecho, sólo una docena de países mostraron mejoras mensurables en el rendimiento científico de sus jóvenes de 15 años, incluidos los sistemas educativos de alto rendimiento, como Singapur y Macao (China), y los de bajo rendimiento, como Perú y Colombia”* (OCDE, 2016. Pág. 3). Sin embargo, en el Anexo A5 del mismo documento se afirma que *“La mejora en el rendimiento medio de ciencias reportada para Colombia se debe casi totalmente a cambios en el enfoque a la escala. El aumento de la puntuación media habría sido de sólo tres puntos (no significativos) si se hubiera utilizado el enfoque de 2015 y la muestra de calibración para escalar los resultados de 2006”* (OCDE, 2016. Pág. 308).

²² Cabe señalar que en las aplicaciones PISA anteriores a 2015, los ítems incompletos al final de la evaluación no se consideraban para estimar los parámetros de cada ítem. Es decir, no se consideraron para establecer las propiedades que asume la curva característica de cada ítem en función de las respuestas, y por tanto su grado de ajuste para el modelo general de estimación de resultados. Pero sí se tomaban en cuenta para calcular los desempeños individuales de cada uno de los estudiantes, una vez definido el modelo general.

Esto quiere decir que esta modificación no abarca a cualquier ítem no respondido, sino a una situación específica: aquella en la que el estudiante “abandonó” la prueba antes de finalizarla, y dejó un conjunto de ítems sin responder al final del cuadernillo o de cada bloque correspondiente a un área. PISA identifica este tipo de ítems no respondidos como *non-reached items* (de aquí en adelante, NRI).

¿Qué implica este cambio? Puede explicarse con un ejemplo sencillo: una prueba estándar de PISA puede tener unas 50 preguntas, que pueden abarcar una o más áreas. Si un estudiante deja las 5 últimas preguntas de la evaluación en blanco (10% del total), estos ítems se considerarían NRI. Hasta la aplicación PISA 2012 inclusive, estos ítems se consideraban erróneos, es decir, el puntaje estimado del estudiante se veía afectado negativamente por estos ítems NRI, obteniendo resultados más bajos de los que se podrían calcular considerando sólo las 45 preguntas anteriores. En cambio, en el 2015 no se consideran administrados: el resultado del estudiante se calcula considerando las 45 preguntas respondidas, y los ítems NRI no afectan negativamente los resultados de un país. En el Anexo III se presenta un ejercicio elaborado por PISA que ilustra cómo cambia el porcentaje de respuestas correctas de un país utilizando estos dos criterios.

En la mayoría de los países, las tasas de no respuesta son muy bajas, por lo que es muy poco frecuente encontrar ítems NRI. Por lo tanto, en estos casos este cambio de criterio no incide sensiblemente en los resultados, y no afecta las comparaciones entre años.

Sin embargo, en el caso de países donde la no respuesta es mayor, puede suceder que haya una mayor cantidad de ítems NRI, lo que impactaría sensiblemente en las comparaciones en el tiempo. Según advierte el informe: “Sin embargo, este nuevo tratamiento de los *non-reached items* podría resultar en puntuaciones más altas de lo que se habría estimado en el pasado para los países con muchos artículos sin respuesta.” (OCDE, 2016).

El siguiente cuadro presenta el porcentaje de ítems NRI de los países evaluados entre PISA 2000 y 2015, considerando el conjunto de ítems aplicados en las tres áreas.

Cuadro 2. Porcentaje promedio de *non-reached items* (NRI) en las evaluaciones PISA (*), por región. Años 2000 a 2015

Regiones	Año 2000 (**)	Año 2003	Año 2006	Año 2009	Año 2012	Año 2015
América Latina (***)	11,9%	9,8%	9,1%	7,5%	6,8%	2,6%
Asia Pacífico	1,7%	1,7%	1,0%	0,9%	0,7%	0,6%
Europa Occidental	3,1%	2,1%	1,2%	1,9%	1,4%	1,2%
Europa Oriental	5,9%	3,2%	1,6%	2,1%	1,5%	1,3%
Países Anglosajones	1,6%	1,3%	1,0%	1,4%	1,3%	0,9%
Países Nórdicos	2,7%	2,4%	1,3%	1,8%	2,0%	1,4%
Promedio países	4,5%	2,8%	2,3%	2,6%	2,2%	1,4%

(*) Se incluye el promedio simple del porcentaje de ítems NRI de cada país, para eliminar el efecto del peso relativo de la población evaluada en cada país. El promedio de países se calculó exclusivamente con aquellos incluidos en las regiones que presenta el cuadro. Se considera el conjunto de ítems aplicados en cada evaluación PISA para lectura, matemática y ciencias.

(**) La aplicación del año 2000 no se controló el efecto de la posición del ítem en el cuestionario (los ítems de lectura siempre se colocaron al inicio). Por lo que la presencia de los *non-reached items* puede variar diferente entre áreas.

(***) Los datos de Argentina 2015 fueron incluidos en el cálculo de América Latina a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC sobre la base de PISA Compendium for the cognitive item responses 2000, 2003, 2006, 2009, 2012 y 2015.

El cuadro muestra información contundente sobre cómo el problema de la existencia de no respuesta al final de las evaluaciones (ítems NRI) se concentró en América Latina ya desde las primeras aplicaciones de PISA.

En la región, el 11,9% de los ítems aplicados en PISA 2000 fueron considerados NRI. Considerando que cada evaluación contenía aproximadamente 60 ítems, este porcentaje representa en promedio 7 ítems por estudiante evaluado. Esta proporción duplica a la de la segunda región con porcentajes más altos, Europa Oriental, con un 5,9% (en promedio, 2,4 ítems NRI por estudiante evaluado).

En las sucesivas aplicaciones, se observa un decrecimiento sostenido en todas las regiones. Para el 2006, el porcentaje de ítems NRI se sitúa por debajo del 2% en las regiones fuera de América Latina, pero allí estos valores se quintuplican, con un 9,8%.

En el período 2006 - 2012, los porcentajes continúan reduciéndose en las aplicaciones de América Latina, llegando al 6,8% en 2012, lo que representa (para una prueba de aproximadamente 56 ítems de longitud) unos 3,8 ítems NRI promedio por estudiante evaluado. En el resto de las regiones, esta proporción se mantiene por debajo del 2%.

En el año 2015, América Latina mejora mucho el porcentaje de ítems NRI en las evaluaciones, reduciéndose a menos de la mitad respecto de 2012. Es importante considerar que en este año la

mayoría de los países participó de PISA respondiendo la evaluación en forma digital, lo que probablemente incida en esta caída²³.

Las bajas tasas de presencia de ítems NRI en otras regiones muestran que el cambio en el criterio de imputación de estos ítems no afecta la comparabilidad en la mayoría de los países evaluados. En América Latina en cambio es evidente que afecta en forma sustantiva la comparación entre aplicaciones, especialmente en los primeros años.

Al interior de América Latina, la situación es diversa entre países y áreas, como puede observarse en el cuadro a continuación.

²³ Si bien no es objetivo de este documento analizar el comportamiento de los ítems NRI en las pruebas en formato digital, los datos globales muestran que hay una diferencia marcada en comparación con los países que completaron la prueba en papel (1,4% de ítems NRI vs 4,4%, respectivamente)

Cuadro 3. Porcentaje promedio de *non-reached items* (NRI) en las evaluaciones PISA (*), por país. Años 2000 a 2015

Países	lectura						matemática					
	Año 2000 (**)	Año 2003	Año 2006	Año 2009	Año 2012	Año 2015	Año 2000 (**)	Año 2003	Año 2006	Año 2009	Año 2012	Año 2015
Argentina (***)	11,7%	N/A	14,5%	11,2%	5,5%	5,3%	14,7%	N/A	12,9%	9,0%	4,6%	4,9%
Brasil	12,7%	11,8%	7,2%	4,9%	6,7%	1,9%	20,4%	10,1%	7,9%	3,9%	5,3%	1,8%
Chile	11,2%	N/A	6,6%	4,1%	3,4%	1,4%	12,1%	N/A	6,3%	2,5%	2,8%	1,4%
Colombia	N/A	N/A	16,6%	11,2%	10,7%	1,9%	N/A	N/A	14,4%	8,8%	9,4%	1,4%
México	3,8%	7,7%	7,4%	6,4%	7,2%	0,6%	6,1%	8,3%	6,5%	4,4%	5,9%	0,7%
Perú	14,9%	N/A	N/A	11,6%	11,5%	0,7%	15,9%	N/A	N/A	9,8%	9,4%	0,6%
Uruguay	N/A	11,4%	11,8%	8,6%	9,1%	2,5%	N/A	10,8%	10,6%	6,1%	7,9%	1,9%

Países	ciencias						Total tres áreas					
	Año 2000 (**)	Año 2003	Año 2006	Año 2009	Año 2012	Año 2015	Año 2000 (**)	Año 2003	Año 2006	Año 2009	Año 2012	Año 2015
Argentina (***)	15,0%	N/A	10,7%	10,6%	4,6%	3,6%	12,8%	N/A	11,8%	10,6%	4,8%	4,6%
Brasil	18,5%	10,0%	6,4%	4,7%	6,6%	2,8%	14,9%	10,4%	6,9%	4,7%	6,0%	2,3%
Chile	12,7%	N/A	5,4%	3,6%	3,3%	2,5%	11,6%	N/A	5,8%	3,7%	3,1%	2,0%
Colombia	N/A	N/A	13,9%	11,3%	10,5%	2,6%	N/A	N/A	14,4%	10,8%	10,0%	2,2%
México	5,7%	8,3%	5,7%	6,3%	7,5%	1,3%	4,5%	8,2%	6,2%	6,0%	6,7%	1,0%
Perú	19,5%	N/A	N/A	11,1%	11,4%	1,2%	15,8%	N/A	N/A	11,1%	10,5%	1,0%
Uruguay	N/A	10,9%	8,1%	7,9%	8,8%	3,0%	N/A	10,9%	9,3%	7,9%	8,4%	2,6%

Nota: se marca en negrita el año en que cada área fue considerada prioritaria en la evaluación

(*) Se considera el conjunto de ítems aplicados en cada evaluación PISA para lectura, matemática y ciencias.

(**) La aplicación del año 2000 no se controló el efecto de la posición del ítem en el cuestionario (los ítems de lectura siempre se colocaron al inicio). Por esta razón, la presencia de los *non-reached items* puede variar diferente entre áreas, y no son estrictamente comparables con el resto de los años.

(***) Los datos de Argentina 2015 fueron incluidos en el cálculo de América Latina a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: CIPPEC sobre la base de PISA Compendium for the cognitive item responses 2000, 2003, 2006, 2009, 2012 y 2015.

El **cuadro 3** permite identificar diferencias marcadas entre países, áreas y aplicaciones, así como significativos cambios a lo largo del tiempo. Esta comparación se centra en los años que comparten la misma área priorizada, de esta forma es posible considerar un universo amplio de ítems de anclaje entre ambos años.

En el año 2000, Perú y Brasil se caracterizan por mostrar los porcentajes más altos, seguidos por Argentina y Chile. En los años siguientes hay una mejora sostenida – expresada en una reducción sistemática de los ítems NRI – en los casos de Brasil y Chile, pero no así en los otros países²⁴.

México, en cambio, muestra porcentajes muy bajos en el 2000 (en el conjunto de las tres áreas, por debajo del 5%), pero no evidencia mejora. Por el contrario, la proporción de ítems NRI se incrementa en 2003, y se mantiene en niveles similares hasta la aplicación del año 2012.

En las aplicaciones 2006 y 2009, pueden identificarse claramente tres grupos diferentes: por un lado, Argentina, Colombia y Perú con valores altos (un porcentaje de NRI que en el conjunto de las tres áreas supera el 10% del total); por otro lado, México, Chile y Brasil (con porcentajes cercanos al 5%); y como tercera categoría el caso de Uruguay, que se sitúa en valores intermedios a ambos grupos.

Para la aplicación de PISA 2012 Argentina reduce sensiblemente el porcentaje de ítems NRI, alcanzado para el conjunto de las tres áreas valores cercanos al 5%, y se posiciona en una situación similar a Brasil y México. Uruguay no mejora, y se acerca a Colombia y Perú que siguen mostrando los porcentajes más altos, en torno al 10%.

Finalmente, en PISA 2015 todos los países – a excepción de Argentina – manifiestan una reducción sustantiva del % de ítems NRI, alcanzando valores cercanos al 2,5%. Es necesario resaltar que Argentina es el único país en la región que no aplicó la evaluación en soporte digital, lo que influye en esta diferencia respecto de los países de la región.

No hay una relación simple que permita calcular cuánto impacta en el resultado de PISA (en puntaje promedio) estas variaciones en los ítems NRI. Los resultados reescalados consideran todos los cambios implementados en la metodología, por lo que no es posible aislar el efecto de los ítems NRI sobre otros aspectos. Lo que queda en evidencia es que aquellos países que muestran mayor porcentaje de ítem NRI (Argentina, Colombia, Perú y Uruguay) son a la vez aquellos que mostraban mayores brechas entre los resultados originales y los nuevos.

Como una forma de aproximarse a esta relación, PISA propone un ejercicio de comparación, calculando cómo varía el porcentaje de respuestas correctas si se consideran o no como incorrectos los ítems NRI. Este ejercicio se presenta en el Anexo III.

²⁴ Para la interpretación correcta de los datos del 2000 por área, es importante tener presente que en la aplicación de ese año los modelos de evaluación no previeron la rotación de la posición de los ítems, incluyeron siempre los de lectura al inicio. Por esta razón, la presencia de los NRI puede variar entre áreas, y no son estrictamente comparables con el resto de los años.

2.3. Comentarios finales sobre el cambio de metodología

El cambio de metodología implementado en PISA 2015 no afecta los resultados de ese año. El problema se suscita en la comparación en el tiempo. Si se hubiese mantenido el mismo criterio utilizado en 2012 – al menos en lo que respecta al tratamiento de los ítems NRI – es muy posible que los resultados de los países de América Latina hubiesen sido muy similares. Como lo muestra el **cuadro 3**, el porcentaje de ítems NRI ha tendido a disminuir en las sucesivas ediciones de la prueba para todos los países de la región, llegando en 2015 a valores muy bajos.

Lo que ha constituido un obstáculo significativo para una comprensión acabada de este problema es el escaso tratamiento de este cambio de metodología en el informe de PISA. Se menciona en un anexo metodológico, mostrando exclusivamente algunos casos extremos, y sin evaluar sus implicancias.

Sin embargo, es destacable que esta modificación permite visibilizar el efecto distorsivo del criterio utilizado en las aplicaciones previas de PISA. Al reescalar los resultados de PISA 2006 – 2012 con la metodología 2015, fue posible identificar el impacto de considerar como erróneos los ítems NRI.

El nuevo criterio utilizado desde 2015 parece ser una forma más adecuada de medir los logros, ya que permite reflejar mejor el desempeño efectivo de los estudiantes (ver Recuadro 2), y por lo tanto, tiene un impacto en la obtención de resultados de mayor calidad técnica. Esto mejorará sustantivamente la utilidad y capacidad explicativa de las aplicaciones de PISA para los países de América Latina en los ciclos futuros.

3. Una reflexión necesaria: ¿qué nos dicen los resultados de las pruebas PISA?

Los cambios metodológicos de PISA abren múltiples preguntas. ¿Qué miden las pruebas PISA? ¿Para qué sirven? ¿Qué conclusiones podemos extraer de sus resultados? Estas preguntas están en el corazón de este estudio y requieren un abordaje específico que conecta cuestiones metodológicas con la gran compuerta epistemológica: ¿qué puede conocerse de un sistema educativo nacional? ¿Pueden medirse sus alteraciones en el tiempo? ¿Con qué precisión? ¿Para qué fines analíticos o prácticos?

3.1. El recorte específico de PISA

Como todas las evaluaciones estandarizadas, PISA tiene una primera limitación crucial: solo mide un recorte muy parcial de los sistemas educativos. Ese recorte comienza por las áreas evaluadas, lengua, matemática y ciencias, que dejan afuera de la prueba numerosos campos de conocimiento y objetivos de los sistemas educativos. Distintos estudios han mostrado el riesgo que implica dar tanta relevancia a PISA hasta el punto de reducir la carga horaria de enseñanza de otras materias como historia, artes o formación ciudadana en el currículum de los países excesivamente preocupados por sus resultados (Pereyra et al., 2011; Meyer, y Benavot, 2013).

También existen limitaciones asociadas a la implementación de una evaluación a gran escala, que impone restricciones al instrumento. Algunas características como el uso exclusivo del soporte escrito, o la priorización de formatos rígidos con preguntas de opciones cerradas, también establecen fronteras sobre lo que puede y no puede conocerse en relación al aprendizaje de la lengua, la matemática o las ciencias.

Otra limitación evidente es que PISA solo mide los aprendizajes de alumnos de 15 años de edad que cursan el 7mo grado en adelante. Esto deja afuera de la prueba a los alumnos de nivel primario y a los alumnos que egresan de la educación secundaria, así como a todos los jóvenes excluidos de la educación a los 15 años de edad. Esta limitación afecta especialmente a los países de América Latina, donde todavía existen niveles importantes de exclusión educativa temprana.

Las pruebas estandarizadas son, además, un recorte muy parcial de los sentidos profundos de un sistema educativo. Cuando se trata de pruebas internacionales se adiciona una capa de neutralidad cultural que impide reconocer las relaciones entre los conocimientos y las culturas de cada contexto y país.

Sin embargo, cabe resaltar los aspectos positivos de la prueba PISA. El primero y principal es que la prueba PISA evalúa competencias, no conocimientos curriculares memorizables. Esto implica tomar como objeto de la evaluación las capacidades de los alumnos de usar el conocimiento en contextos diversos. Los **cuadros 4, 5 y 6** muestran qué significa alcanzar el nivel 6 (máximo), el nivel 2 (considerado un umbral de conocimientos básicos) y los que están debajo del 2 en PISA en cada área evaluada.

Cuadro 4. Niveles de desempeño en lectura en PISA

Nivel	¿Qué pueden hacer los estudiantes de cada nivel?
6	<p>En el nivel 6, los estudiantes pueden inferir, comparar y hacer contrastes con precisión y detalle. Asimismo, pueden comprender completa y detalladamente uno o más textos, e incluso integrar información de más de un texto. Estarán capacitados para comprender ideas nuevas- aun cuando hay mucha información que compite con estas ideas- y generar categorías abstractas para posteriores interpretaciones. También, pueden evaluar críticamente textos complejos sobre temas poco familiares o elaborar hipótesis a partir de estos textos. Para ellos, se considera diversos criterios y perspectivas posibles, y se aplica saberes previos complejos. En este nivel, los alumnos pueden hacer análisis precisos y atender a detalles del texto que suelen pasar desapercibidos.</p>
2	<p>En el nivel 2, los estudiantes pueden ubicar uno o más datos, que podrían tener que ser inferidos y cumplir con varias condiciones. Asimismo, son capaces de realizar comparaciones o contrastes basados en un solo criterio. Además, pueden reconocer la idea principal de un texto, comprender relaciones y construir significado a partir de una parte del texto, cuando la información no es notoria y las inferencias a realizar son de baja demanda. Por último, están capacitados para comparar o conectar al texto con saberes previos ajenos al texto, recurriendo a su experiencia personal.</p>
1a	<p>En el nivel 1a, los estudiantes pueden ubicar uno o más datos independientes expresados explícitamente, reconocer el tema central o el propósito del autor en textos sobre temas conocidos, y establecer las relaciones sencillas entre información del texto y saberes de la vida cotidiana. De igual modo, localizan datos notorios en el texto cuando hay poca o ninguna información que compite con estos. En este nivel, las preguntas orientan de manera explícita a los estudiantes para que tomen en cuenta los factores relevantes de la tarea y del texto</p>
1b	<p>En el nivel 1b, los estudiantes pueden ubicar un solo dato específico y notorio de un texto breve y sintácticamente sencillo, cuando el contexto y el tipo de textos son familiares (por ejemplo, narraciones o listados simples). Además, establecen relaciones sencillas entre información contigua. Cabe anotar que los textos propios de este nivel tienen muy poca información en competencia: además, presentan diversas ayudas al lector, como repetición de información o empleo de imágenes o símbolos conocidos.</p>

Fuente: OCDE (2013b)

Cuadro 5. Niveles de desempeño en matemática en PISA

Nivel	¿Qué pueden hacer los estudiantes de cada nivel?
6	En el nivel 6, los estudiantes pueden conceptualizar, generalizar y utilizar la información sobre la base de sus investigaciones y modelos de situaciones de problemas complejos. Asimismo, pueden relacionar diferentes fuentes de información y tipos de representaciones, y pasar de una a otra con flexibilidad. Los alumnos de este nivel, serán capaces de pensar y razonar con matemática avanzada. Pueden aplicar su conocimiento, comprensión e intuición, así como su dominio de las operaciones y relaciones matemáticas simbólicas y formales, para desarrollar nuevos planteamientos y estrategias frente a situaciones nuevas. Del mismo modo, podrán formular y comunicar con precisión sus acciones y reflexiones referidas a sus resultados, interpretaciones y argumentos, y su pertinencia a situaciones originales.
2	En el nivel 2, los estudiantes pueden interpretar y reconocer situaciones en contextos que requieren de una inferencia directa. De igual modo, pueden extraer información relevante a partir de una única fuente y hacer uso de un único modo de representación. A su vez, podrán utilizar algoritmos, fórmulas, procedimientos o convenciones básicos. También, son capaces de realizar razonamientos directos e interpretaciones literales de sus resultados.
1	En el nivel 1, los estudiantes pueden responder a las preguntas que involucran contextos conocidos en los que se encuentra toda la información necesaria y las preguntas están claramente definidas.

Fuente: OCDE (2013b).

Cuadro 6. Niveles de desempeño en ciencia en PISA

Nivel	¿Qué pueden hacer los estudiantes de cada nivel?
6	En el nivel 6, los estudiantes, de forma consistente, pueden identificar, explicar y aplicar su conocimiento científico y su conocimiento acerca de la ciencia en variadas y complejas situaciones de la vida real. Asimismo, pueden relacionar diferentes fuentes de información y explicaciones y usan evidencia de esas fuentes para justificar sus decisiones. Clara y consistentemente pueden demostrar un pensamiento y razonamiento crítico avanzado y están dispuestos a servirse de esa comprensión científica para sustentar las soluciones planteadas en situaciones desconocidas en los ámbitos científicos y tecnológicos. Los estudiantes en este nivel pueden usar su conocimiento científico y desarrollar argumentos que apoyen recomendaciones y decisiones centradas en situaciones personales, sociales y globales.
2	En el nivel 2, los estudiantes tienen el conocimiento científico adecuado para proporcionar posibles explicaciones en contextos familiares o para extraer conclusiones basadas en investigaciones simples. Además, pueden razonar directamente y hacer interpretaciones literales de los resultados de la indagación científica o de la solución de problemas tecnológicos.
1	En el nivel 1, los estudiantes tienen un conocimiento científico tan limitado que solo pueden aplicarlo a escasas situaciones familiares. Pueden ofrecer explicaciones científicas que son obvias y deducibles explícitamente de las evidencias dadas.

Fuente: OCDE (2013b).

Evaluar competencias implica concebir al conocimiento como un conjunto de disposiciones y capacidades de acción. Más concretamente, implica traducir el conocimiento en poder. Solo si los alumnos aprenden de manera tal que puedan ser capaces de usar el conocimiento para salir de su destino social se cumplirá el derecho pleno a la educación. La prueba PISA es uno de los instrumentos más rigurosos y pertinentes para medir la capacidad de empoderamiento de los alumnos a partir de su tránsito escolar.

Aquí tenemos una paradoja y un malentendido. La paradoja es que PISA limita y expande al mismo tiempo las concepciones curriculares de los países. Cuanto más importante es PISA mayor peso tiene esta paradoja. Al mismo tiempo que puede producir un efecto de vaciamiento de saberes fundamentales no evaluados también puede promover replanteos que vayan de conocimientos fragmentarios y enciclopédicos a capacidades de actuar y usar el conocimiento para cambiar el destino social de los alumnos.

Se trata de una paradoja que debe ser resuelta con madurez por las políticas curriculares de los países. Los riesgos deben estar claros para las autoridades políticas. El borramiento de los contextos culturales del currículum o de áreas enteras (como las artes, las ciencias sociales, la ciudadanía, la educación física, etc.) es un peligro que debe ser evitado. Así como también es necesario realizar adaptaciones de los enfoques por competencias adecuadas en cada contexto, sin imponer cambios inviables o desconocedores de los puntos de partida, saberes y opiniones de los colectivos docentes.

Más allá de esta paradoja, hay un malentendido que rodea a la prueba PISA. Se trata de aquellas posiciones críticas de la prueba por su supuesto rol de colonización cultural para favorecer la globalización económica al servicio del mercado. El hecho que la prueba sea administrada por la OCDE aumenta estas sospechas, que se traducen en una resistencia de ciertos sectores políticos.

Esta visión crítica de la prueba PISA tiene elementos fundados en sus argumentos. Como hemos señalado, PISA elude otras áreas curriculares, suprime contextos culturales y evita juzgar qué sentido social, político y ético adquieren los conocimientos que evalúa. Dicho esto, la visión curricular de PISA, centrada en las capacidades de usar el conocimiento en diferentes contextos, parece mucho más adecuada para defender el derecho pleno a la educación que otras versiones curriculares más tradicionales, todavía predominantes en la mayoría de los países de América Latina. Habría que ser mucho más críticos con los diseños curriculares, las formas de enseñanza predominantes y los aprendizajes adquiridos en los sistemas educativos oficiales de los países, antes de ser tan críticos frente a PISA.

Enseñar “para PISA” implica enseñar capacidades profundas que permiten ampliar los horizontes culturales, sociales y políticos de los alumnos. Un alumno en nivel 5 o 6 de PISA es un sujeto plenamente capaz de afectar su mundo y el de los demás. El problema central no es lo que mide PISA sino las limitaciones pedagógicas, culturales, presupuestarias y políticas de los sistemas educativos que enseñan, mayormente, a repetir conocimientos de memoria y no asocian el saber con el poder de cambiar el propio destino y el de las sociedades donde habitan los estudiantes.

Es cierto que los saberes de PISA favorecen la globalización económica con sujetos más competentes para el mercado. Pero no lo hacen de forma lineal: lo que evalúa la prueba es tan profundo y multidimensional que puede ser utilizado para múltiples propósitos. Que un alumno sea capaz de comprender un texto, relacionarlo con otros y tomar posición propia fundamentada no es un saber “útil para el mercado”; es, antes que nada, un saber útil para ser dueño de su propio destino.

PISA simplifica, descontextualiza, sesga el currículum, evita los aspectos culturales del conocimiento y desliga al saber de sus dimensiones ética, política y social, pero es una prueba superadora del saber curricular y pedagógico dominante en la mayoría de los países y en especial en América Latina. Preparar para PISA, con todas estas limitaciones, es aumentar el piso del derecho a la educación, es llenar de poder los conocimientos de los alumnos, es transformar cognitivamente el sentido de los aprendizajes. Pero, como toda arma de doble filo, debe ser utilizada con extremo cuidado y precisión.

3.2. La validez de PISA

Hecho este primer argumento sobre el sentido curricular de las pruebas PISA, cabe entonces abordar las preguntas sobre qué tipos de usos permite la metodología de las pruebas PISA²⁵. ¿Qué tipo de interpretaciones y usos pueden realizarse de las pruebas? ¿Hasta dónde llega su validez?

Los cambios metodológicos de la edición 2015 dejan al desnudo dos aspectos críticos sobre la validez de las pruebas PISA para medir la evolución de la calidad educativa en los sistemas educativos a través del tiempo. El primero refiere a la manera en que PISA presenta y comunica los resultados. El segundo es un problema de implementación que resulta mucho más relevante para los países participantes.

El problema de comunicación de los resultados 2015 por parte de la OCDE refiere a la poca visibilidad con que fueron presentados los cambios metodológicos analizados en el apartado 2. Los cambios afectaron la comparabilidad intertemporal de los resultados, especialmente en algunos países de América Latina, lo cual debió ser aclarado en el cuerpo central de los informes de PISA y en los informes por países. Los casos de Colombia y Uruguay muestran las implicancias de este problema: las autoridades de ambos países presentaron las mejoras en la calidad educativa medidas por PISA en base a los informes oficiales, cuando en realidad, corrigiendo los resultados para hacerlos estrictamente comparables, esas mejoras no existieron. En vez de presentar estas correcciones en los anexos metodológicos, PISA podría haber aclarado el problema en el cuerpo central de sus informes para evitar malentendidos.

El segundo problema es más importante y refiere a la manera en que PISA es implementada por los países. Como observamos en el **cuadro 3** del capítulo 2, los países de América Latina tenían altas tasas de pruebas que no eran completadas por los alumnos, dejando preguntas sin responder (*non-reached items*). Por ejemplo, Brasil tuvo un 20,4% de ítems no respondidos en la prueba de matemática del año 2000 y apenas un 1,8% en 2015. Perú tuvo un 19,5% de ítems no respondidos en la prueba de ciencia en 2000 y apenas un 1,2% en 2015. En todos los países de América Latina se comprobó que, con el paso de las aplicaciones de PISA, se logró un gran avance en la tasa de respuesta completa de las pruebas. Esto puede ser visto como el resultado de dos procesos combinados:

1. La adaptación cultural a la aplicación de pruebas. Dado que la mayoría de los países de la región comenzó a implementar evaluaciones nacionales censales recientemente, quizás esto haya ayudado a crear una cultura de la respuesta al ejercicio específico de una prueba estandarizada. La adaptación no implica solamente entender el tipo de prueba que tienen en

²⁵ Entre los numerosos estudios sobre aspectos metodológicos de la prueba PISA, véase en particular la metaevaluación de Fernández-Cano (2016), que sintetiza varias cuestiones.

sus manos los alumnos (que es todo un proceso en sí mismo y lleva tiempo y preparación), sino el hecho de tener que responder una prueba que no tendrá ninguna consecuencia individual para ellos y realizar el máximo esfuerzo. Los alumnos pueden estar acostumbrados a realizar pruebas; después de todo, es una práctica común en los sistemas educativos. Pero otra cosa es hacer una prueba que no mide su aprendizaje ni tiene consecuencias sobre ellos, sino que mide un sistema educativo completo. Esto implica construir a gran escala la disposición de hacer el mayor esfuerzo para un fin relativamente abstracto: el mejor logro del sistema educativo nacional.

2. La reparación específica para completar la prueba²⁶. Tan o más importante que la práctica de otras evaluaciones estandarizadas de cada país parece haber sido la preparación directa para la prueba PISA que la mayoría de los países de la región comenzó a implementar a medida que pasaron las ediciones de la prueba. Esto se vincula intrínsecamente con la importancia que fue cobrando PISA en los países²⁷. Los responsables políticos de la conducción de los sistemas educativos nacionales comenzaron a sentir una creciente presión por los resultados que se tradujo en distintas estrategias para preparar a los alumnos para hacer su mayor esfuerzo en las pruebas PISA.

Sobre este segundo punto es importante considerar que PISA envía la lista de la muestra de escuelas seleccionadas con aproximadamente 6 meses de anticipación. Dentro de ese período de aviso y la toma de la prueba los países pueden realizar un proceso de preparación para reforzar el compromiso de cada escuela, sus docentes y alumnos, con la aplicación de la prueba. Esto no es ilegal, al contrario, es parte del protocolo esperado.

²⁶ Este tema, de gran complejidad e importancia, ha sido escasamente estudiado por la comunidad internacional de especialistas. En la amplia revisión bibliográfica realizada para este estudio, pocos trabajos lo han estudiado en detalle (Solano-Flores y Milbourn, 2016; Wuttke, 2007; Joodouin, 2014), y ninguno que ha realizado algún tipo de trabajo empírico, por ejemplo, verificando en más de un país cómo se implementó la prueba PISA en el territorio.

²⁷ Solano-Flores y Milbourn (2016) introducen el término de *capacidad evaluativa* para referirse al grado de desarrollo de las capacidades técnicas de las instituciones públicas de los estados nacionales para la implementación de evaluaciones de desempeño a escala. Particularmente, hacen referencia a tres aspectos claves: (1.) el grado en el que un país ha desarrollado o es capaz de desarrollar y utilizar instrumentos de evaluación técnicamente apropiados; (2.) el grado en el que un país ha creado o es capaz de crear y mantener un sistema evaluativo, y (3.) el grado en el que un país es capaz de utilizar la información obtenida mediante aquellos instrumentos y sistemas de evaluación para informar sus políticas y prácticas.

En la medida en que los países van desarrollando su capacidad evaluativa, va mejorando la calidad de los resultados obtenidos. Al respecto, es interesante referir al debate generado en algunos países sobre las acciones de “sensibilización” que algunos ministerios desarrollaron en las escuelas a ser evaluadas por PISA. Con este término se hace referencia a la preparación de las escuelas y los estudiantes para la recepción de la evaluación PISA. Las buenas prácticas de evaluación sugieren la implementación de estas acciones de preparación. No se trata de entrenar a los estudiantes en los saberes y capacidades evaluados por PISA para que obtengan mejores resultados. Por el contrario, lo que se busca es favorecer la construcción de un entorno adecuado para que la prueba pueda realizarse sin dificultades, y prever que no haya factores distorsivos de los resultados. Por ejemplo, si los estudiantes no están acostumbrados a responder una prueba basada en *múltiple choice*, es importante que se familiaricen con el formato para que no sea en sí mismo una restricción que les dificulte expresar sus saberes.

Estas acciones de sensibilización tienen efectos positivos y negativos: por un lado, contribuyen a obtener información de mejor calidad. Pero también introducen sesgos en la comparación en el tiempo con aplicaciones donde no se preparó a las escuelas para recibir a PISA. Se corre el riesgo de interpretar como mejoras en el aprendizaje resultados que reflejan simplemente mejores condiciones de implementación.

Por lo tanto, la decisión política de generar un contexto favorable para la implementación de PISA tiene que considerar estas situaciones, y los países deben tener presente al momento de analizar los resultados en el tiempo el impacto posible de estas acciones.

Lo que se busca con estas acciones es favorecer la captura de información de calidad, evitando la introducción de sesgos en los resultados. Las respuestas a la evaluación deben reflejar lo más fielmente posible las diferencias en los aprendizajes de los estudiantes, sin intervención de otros factores. Para ello, estas prácticas de preparación de las escuelas se proponen, por ejemplo:

- Suplir la ausencia de incentivos o motivaciones concretas que los alumnos tienen para contestar la prueba con su mayor esfuerzo. Dado que la prueba dura dos horas y demanda un muy alto nivel de concentración y esfuerzo cognitivo, esta preparación es fundamental para que los alumnos realmente se dispongan a demostrar todo lo que saben.
- Familiarizar a los estudiantes con el formato de la evaluación. Puede ocurrir que los estudiantes no estén acostumbrados a responder el tipo de consignas que propone PISA, si la escuela no incluye este tipo de pruebas en sus evaluaciones. Esta falta de familiaridad puede constituir una restricción que no se vincula con los aprendizajes sino con el manejo de la herramienta.
- Sugerir estrategias de gestión de la evaluación e incentivar a la elección de una respuesta. Como fue trabajado en el capítulo 2, la calidad los resultados de PISA se asocian a las mayores tasas de respuesta (correctas o incorrectas, indistintamente). Las diferentes estrategias de los estudiantes frente a preguntas que demandan tiempo, la tendencia a arriesgar respuestas o a dejar preguntas en blanco, las reacciones frente a la falta de tiempo para completarla, inciden en los resultados sin necesariamente implicar diferencias en los conocimientos.
- Preparar a las escuelas para facilitar la reorganización de los grupos frente a la aplicación de PISA. La elección de estudiantes de 15 años implica tener que dividir y reagrupar los grupos clase, entre diferentes secciones y grados, prever espacios para el grupo de evaluados, y personal docente para acompañar el proceso.

En América Latina existe una indicación indirecta concreta de que el esfuerzo por realizar la prueba aumentó de manera constante entre la primera edición de 2000 y la última de 2015: el porcentaje de ítems no alcanzados disminuyó de 12,8% a 4,6%. La hipótesis que puede establecerse es la siguiente: la disminución de la no respuesta a preguntas de la prueba es una expresión del mayor esfuerzo que realizaron los alumnos tomando conciencia de la importancia de contestar con su máxima aptitud la prueba PISA. Ese mismo mayor esfuerzo pudo haberse traducido en mejores resultados en el resto de la prueba, algo que no podemos medir de manera directa, pero que podemos inferir en la gran disminución de las preguntas no contestadas (*non-reached items*) en los países de la región.

Recuadro 2. Los ítems no alcanzados y los aprendizajes: ¿es pertinente el cambio metodológico?

La imputación de ítems NRI a respuestas incorrectas, criterio implementado por PISA hasta el año 2012, se sustenta en el supuesto de que pruebas incompletas expresan exclusivamente dificultades en el desempeño, asociadas a la falta de tiempo para completarla. Por lo tanto, una estimación “justa” de los resultados de PISA debiera penalizar a los países donde estos problemas se manifiestan con mayor frecuencia. Esta asociación entre pruebas incompletas y aprendizajes más deficientes puede ser puesta en cuestión a partir de diferentes argumentos.

Por ejemplo, Wuttke (2007) analiza cómo reaccionan los estudiantes cuando se están quedando sin tiempo en una evaluación. El autor reconoce, por ejemplo, que los estudiantes holandeses priorizan responder todo, bien o mal, y recurren más a adivinar las respuestas. Los austríacos y alemanes saltean los ítems más difíciles desde el inicio, lo que les deja suficiente tiempo para terminar la prueba sin acelerar mucho su ritmo. La presión del tiempo “tomaría por sorpresa” a los estudiantes

griegos, que muestran muy altas tasas de respuestas correctas al inicio y muchos ítems sin responder al final. Estos se verían perjudicados por razones que no tienen que ver con la calidad de los aprendizajes: la estrategia adoptada trae como consecuencia una mayor presencia de ítems incompletos al final de la prueba (ítems NRI).

Un interesante estudio desarrollado en Canadá (Jodouin, 2014) explora los factores que inciden en las tasas de no respuesta en los diferentes contextos culturales y lingüísticos del país. Entre sus hallazgos, el autor reconoce la importancia de la existencia de lo que denomina una “cultura de la evaluación”, un factor fuertemente asociado a la omisión de respuestas. Entiende como tal a la presencia de políticas educativas orientadas a la evaluación, la inclusión de estas temáticas en la formación docente, la frecuencia de aplicación de evaluaciones estandarizadas y las políticas de uso de los resultados. Estos son aspectos que están fuertemente asociados a la decisión de un estudiante de no responder un ítem, arriesgar una respuesta, o abandonar la prueba antes de finalizarla.

En este segundo ejemplo, la presencia de ítems NRI en las pruebas tampoco estaría asociada a los niveles de aprendizaje de los estudiantes. En los contextos donde no se haya instalado una cultura de la evaluación, los resultados de los estudiantes se verán más afectados.

La incorporación de las rutinas de evaluación, la sensibilización a las escuelas y estudiantes sobre su importancia, una mayor preparación de los equipos técnicos, son aspectos que favorecen a la obtención de respuestas de mayor calidad. Pero mayor calidad en la información no debe confundirse con un mejor desempeño de los estudiantes, ya que se trata de dos dimensiones completamente diferentes. El hecho de que los estudiantes completen más ítems no significa que lo hagan mejor, que sus respuestas sean más certeras.

En la mayoría de los países de América Latina, las políticas de evaluación educativa comenzaron a tomar forma en la década de los '90, y terminaron de consolidarse en el nuevo siglo (Rivas, 2015). Por lo tanto, en numerosos países, la participación en PISA se inició en un marco temporal muy cercano a la instalación de una evaluación nacional. Los sistemas educativos no tenían incorporada la evaluación estandarizada como parte de los relevamientos regulares, y no formaba parte de su cultura institucional.

Es necesario profundizar el análisis de la omisión de ítems en América Latina para identificar con certeza las razones por las cuáles es tan significativo el porcentaje de ítems NRI en las aplicaciones iniciales de PISA, y a qué factores responden. A priori, la evidencia disponible, aunque parcial y no concluyente, interpela la consideración de los ítems NRI como respuestas erróneas. Bajo este tratamiento de los NRI, la región se habría visto perjudicada en las primeras aplicaciones de PISA por la escasa penetración de políticas nacionales de evaluación, y a su vez habría sido “premiada” en los ciclos siguientes con una mejora de los desempeños más asociada a la incorporación de la cultura de la evaluación que a la mayor calidad de los aprendizajes. En este sentido, el cambio implementado en 2015 estaría yendo en la dirección correcta al contribuir a eliminar el efecto de factores ajenos al aprendizaje en los resultados.

Esto nos lleva a una situación paradójica: mientras se logró mejorar la calidad de la implementación de la prueba, aumentando el compromiso de los alumnos con responderla en plenitud de sus facultades cognitivas, se pudo haber afectado la comparabilidad estricta de los resultados.

Más allá del problema ya documentado en el apartado 2 sobre el cambio metodológico, aquí estamos señalando un aspecto que no puede ser documentado por la propia prueba PISA: ¿cambió

el nivel promedio de esfuerzo de los alumnos al responder las pruebas PISA en sus sucesivas ediciones? Si cambió el nivel de esfuerzo de los alumnos al contestar la prueba, ¿cómo podemos saber si realmente mejoraron (o no empeoraron) sus aprendizajes o si solo mejoró su predisposición a responder la prueba con su máximo esfuerzo?

No hay forma de contestar esta pregunta con la investigación realizada. Esto requeriría un trabajo cualitativo de entrevistas a diversos actores involucrados en la implementación de las pruebas PISA en sus sucesivas ediciones para reconstruir parcialmente los mecanismos de preparación y esfuerzo de los alumnos al contestar los ejercicios.

La imposibilidad de contestar esta pregunta debe ser parte del abordaje de la validez de los resultados de las pruebas PISA en la comparación inter-temporal de cada país. El mayor peso político de la prueba en los países posiblemente haya causado una mayor preocupación por preparar a las escuelas que pudo haber tenido un correlato en la pérdida parcial de la comparabilidad de los resultados.

Lo positivo de este aumento del compromiso con la realización de la prueba es que las comparaciones hacia adelante serán más fiables si se mantiene ese nivel de preparación. Si antes no podíamos saber en qué proporción los resultados de aprendizaje se debían a una falta de esfuerzo y motivación de los alumnos para contestar la prueba (una dimensión muy variable según países y culturas), ahora quizás podamos tener una medida más clara de sus logros reales de aprendizaje, dado que parece haber sido más equilibrada la motivación para contestar la prueba en América Latina. Esta es una pregunta abierta, dado que solo tenemos un indicio empírico concreto: la disminución de la tasa de ítems no contestados (*non-reached items*) por los alumnos.

Hechas estas aclaraciones, se pueden tomar todos estos elementos y los analizados en el apartado anterior para sintetizar la validez de las pruebas PISA, al menos en las ediciones 2000-2015.

1. *La comparación de los resultados de los países es válida, especialmente si se observan varias ediciones de las pruebas PISA y se hace énfasis en cada área de aprendizaje prioritaria evaluada.*

PISA tiene problemas de comparabilidad intercultural que ya han sido estudiados y que pueden sesgar el diseño de las preguntas de la prueba y la implementación, dada la variedad de contextos en los cuales se implementa²⁸. Si bien este punto limita la comparabilidad de los resultados entre países, el diseño de la prueba es robusto y riguroso para permitir justamente sobrepasar estas limitaciones.

De todas formas, se recomienda analizar el resultado comparado de los países el área evaluada prioritariamente en la edición en cuestión²⁹.

²⁸ Diversos autores han analizado críticamente los problemas de comparabilidad cultural de las pruebas PISA (Meyer y Benavot, 2013; Goldstein, 2004; Carnoy, 2015; Kreiner, 2011). Otros autores han cuestionado la comparabilidad resaltando aspectos específicos como la validez del índice de nivel socioeconómico y cultural (Carnoy y Rothstein, 2013; Jerrim, 2012), la traducción de las pruebas a distintos idiomas (Eivers, 2008; Puchhammer, 2007 ¿cuál es?), los problemas de la tasa de no respuesta de las pruebas (Jerrim, 2013; Wuttke, 2007) o las diferencias temporales de aplicación de la prueba según el calendario escolar de cada país (O'Leary, 2001; Prais, 2003).

²⁹ PISA coloca énfasis en una de las tres áreas cada nueve años, y eso permite una comparación más robusta de los resultados de los distintos países. En 2000, el área priorizada fue Lectura; en 2003, se priorizó Matemática; y en 2006, se priorizó Ciencias. En 2009 se volvió a colocar el foco en Lectura, volviendo a iniciar el ciclo (con Matemática como foco de la edición 2012 y Ciencias como foco de 2015).

La prueba PISA también tiene dificultades para su adaptación a los países de América Latina, dado que fue diseñada en el contexto de la OCDE y muchos aspectos propios de la región y de otros países menos desarrollados que participan como invitados son tratados de manera más marginal. En cierta medida esto se evidencia en lo analizado a lo largo del capítulo 2. Algunos documentos específicos se han elaborado para analizar las particularidades de Iberoamérica en sus distintas participaciones en la prueba PISA (OCDE, 2016).

2. *La evolución de los resultados por país a lo largo del tiempo es válida pero se recomienda tomar recaudos para sacar conclusiones apresuradas sobre los cambios producidos de una edición a otra*³⁰.

PISA también asegura una serie de controles metodológicos a partir de ítems de anclaje para comparar los resultados a lo largo del tiempo. Sin embargo, cabe resaltar que cada edición prioriza un área. Esto implica que la comparación en el tiempo tendrá mayor validez entre las ediciones que priorizaron la misma área³¹.

PISA indica que las comparaciones son válidas desde el año de anclaje inicial hacia adelante en todos los casos: lengua desde 2000, matemática desde 2003 y ciencia desde 2006. Esto permite comparaciones de una edición a otra en todas las áreas a partir de 2006. Sin embargo, debería ser muy cauto el análisis de una edición a la siguiente a la luz de lo analizado hasta aquí especialmente para América Latina. Las condiciones de aplicación de la prueba pueden variar de una edición a otra, especialmente el nivel de esfuerzo que realizan los alumnos para responder una tarea tan demandante.

Por eso se recomienda observar trayectorias más prolongadas y no extraer conclusiones apresuradas sobre los avances o retrocesos de los países. Más aún: se sugiere analizar los resultados de las pruebas PISA combinados con los de otras evaluaciones de la calidad educativa, especialmente las que realiza cada país³². Esto permitirá lograr análisis más robustos sobre la evolución de los sistemas educativos.

3. *El seguimiento de los resultados en las pruebas PISA a lo largo del tiempo no es una evidencia suficiente para extraer conclusiones apresuradas sobre la mejora de ciertos países, y menos aún para tomar lecciones de política de ellos.*

No es conveniente tomar las pruebas PISA para realizar recomendaciones de política educativa basadas en la trayectoria de los países en el tiempo. Existen dos argumentos muy distintos para realizar esta advertencia. El primero se vincula con lo señalado previamente: en especial los países de América Latina parecen mostrar esfuerzos diferentes en el compromiso de los alumnos para responder las preguntas en cada edición de la prueba PISA. Observar la evolución de los resultados requiere cuidados y lecturas complementarias de otras evaluaciones para no asumir un supuesto avance o retroceso de los países en períodos cortos o situaciones irregulares a lo largo de los años. Es más probable que haya variado la forma de aplicación de la prueba que encontrar en los sistemas educativos nacionales grandes alteraciones de sus resultados cada tres años. En cambio, si la

³⁰ Distintos autores han cuestionado la validez de la comparabilidad intertemporal de los resultados de las pruebas PISA, en casos como Portugal (Freitas et al., 2015), Estados Unidos (Carnoy, 2013), Inglaterra (Prais, 2003), Corea del Sur (Eivers, 2010) y Brasil (Klein, 2011).

³¹ Es decir, Lectura entre 2000 y 2009, en Matemática entre 2003 y 2012 y en Ciencia entre 2006 y 2015.

³² Carnoy (2015) remarca que existen diferencias considerables en la trayectoria intertemporal de los resultados de los países en las pruebas PISA y TIMMS, que cuestionan su validez como medida objetiva de un sistema educativo nacional.

tendencia se mantiene en varias ediciones y se corrobora con otras evaluaciones, entonces pueden elaborarse hipótesis más confiables sobre la trayectoria de los aprendizajes de un sistema educativo.

Un segundo argumento remite a la validez de las fuentes para realizar recomendaciones de política educativa. Para este fin se sugiere analizar los contextos de cada país y la historia de sus sistemas educativos antes de realizar recomendaciones aisladas en base a una mejora durante un período dado. Algunos países pueden mejorar sus resultados en base a los logros del crecimiento económico y la mejora social de su población, otros por mecanismos de extrema presión sobre los alumnos, al punto de ahogar su infancia y su adolescencia en el “infierno de los exámenes”. Hemos estudiado el caso de Perú (Rivas 2015) representativo de la primera imagen y diversos estudios han analizado el caso de Corea del Sur (Seth, 2007), representativo de la segunda. También podríamos citar casos de países que han priorizado la reconstrucción de su tejido social a partir del énfasis en la educación para la paz y la ciudadanía, a los cuales no se les podría imponer el mismo balance de prioridades curriculares que a otros en relación a la enseñanza de lengua, matemática y ciencia. Colombia podría ser un buen ejemplo de esto.

Recomendar políticas educativas requiere un balance de evidencias empíricas con estudios más profundos sobre los contextos culturales, las condiciones económicas y sociales y visiones de la educación en términos políticos, filosóficos, pedagógicos, curriculares y económicos.

4. *Los cuestionarios complementarios de PISA permiten acceder a información valiosa sobre los sistemas educativos.*

La información que recolecta PISA a través de los cuestionarios a directores y estudiantes (y a otros actores del sistema educativo en forma optativa) constituye un recurso de suma importancia para el diagnóstico de la situación de los países. Los datos que sistematiza permiten “bucear” dentro de los promedios nacionales para construir información clave para alimentar un enfoque de justicia educativa: permiten identificar indicios de factores que se asocian al aprendizaje, reconocer poblaciones vulnerables, y dimensionar la magnitud de las disparidades y los cambios que pueden ocurrir en el tiempo.

La implementación de modelos estadísticos multinivel a partir de estos datos contribuye a reconocer la influencia específica de ciertos factores (como la enseñanza) controlando la incidencia de otros (como el nivel socioeconómico), rompiendo con las miradas deterministas que reducen las explicaciones a unas pocas variables. Aunque es necesario prevenir también en estos modelos la tentación simplista de olvidar el alcance epistemológico de un análisis que no considera los factores no reductibles a métricas.

La información que producen los cuestionarios complementarios no solamente es útil en su vinculación con los resultados de PISA. Es una valiosa herramienta de caracterización de los estudiantes y las escuelas. Especialmente en países con bajo desarrollo de sistemas nacionales de información, estos datos iluminan aspectos de los sistemas educativos que son desconocidos, y le otorga a la opinión de los estudiantes un lugar central para entender qué ocurre en las escuelas.

Tomando estas advertencias, y reforzando al mismo tiempo la importancia de contar con las pruebas PISA como un instrumento necesario y útil para ayudarnos a comprender la complejidad de los sistemas educativos, el apartado siguiente aborda la pregunta de investigación país por país. ¿Cómo evolucionaron los logros de aprendizaje en América Latina? Para ello se analizan no solo los resultados en PISA sino en otras evaluaciones internacionales y nacionales, así como diversos indicadores del contexto social, económico y educativo.

4. Resultados medidos por evaluaciones de la calidad por países

Este apartado presenta la descripción y el análisis de los resultados de Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay en las evaluaciones de la calidad educativa implementadas entre los años 2000 y 2016, junto con referencias sintéticas a su situación en torno a otros indicadores educativos y sociales³³. La ubicación de este capítulo a esta altura del documento no es casual: solo después del desarrollo de ciertas alertas y claves de lectura e interpretación se invita al lector a recorrer la historia de los resultados educativos medidos por evaluaciones de calidad en América Latina. Se trata de un recorrido marcado por múltiples interrogantes sobre los resultados, las políticas educativas y sus contextos. En el camino, el análisis intenta ofrecer respuestas parciales a estos interrogantes y abrir nuevas preguntas para futuras investigaciones.

Los resultados se presentan para los siete países abordados en el estudio original (Rivas, 2015), tanto en lo que respecta al nivel secundario (medido por PISA y evaluaciones nacionales) como al nivel primario (medido por las evaluaciones del LLECE-UNESCO y evaluaciones nacionales). En torno a los resultados de PISA, se los presenta tanto en términos de lo publicado por la OCDE en el informe central como de los resultados del ejercicio de reescalamiento tras los cambios metodológicos implementados en 2015. Esta información y el análisis que de ella se deriva es complementada por una síntesis del comportamiento de los principales indicadores educativos y sociales en el período, para favorecer la lectura de los resultados en relación al contexto propio de cada país.

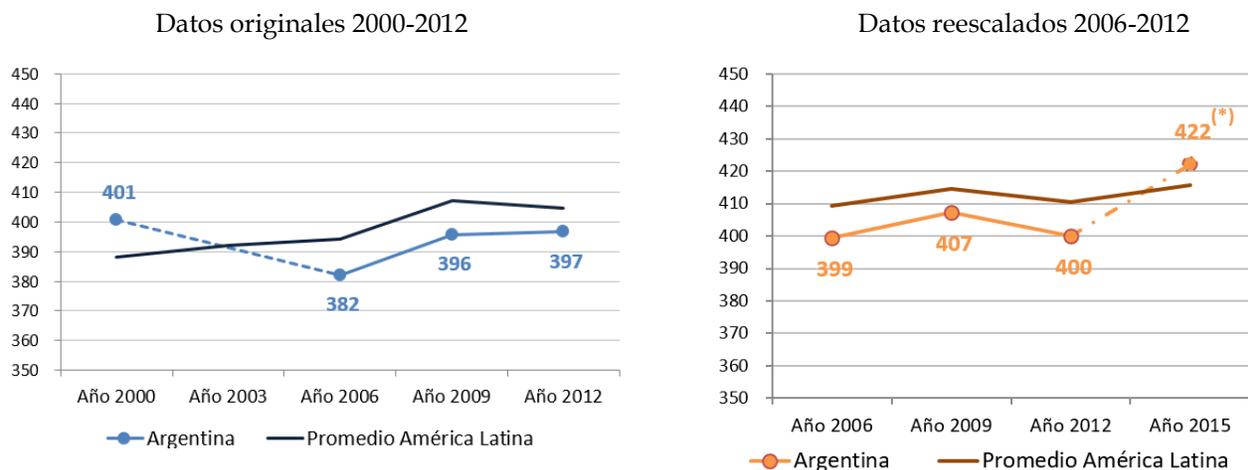
4.1. Argentina

Resultados comparados de las evaluaciones de la calidad educativa

Argentina participó en la primera aplicación de las pruebas PISA en el año 2000, y a partir del 2006 fue parte de la evaluación en forma continua hasta el 2015. En este último año, la implementación fue afectada por problemas en el marco muestral, y fue excluida de la publicación, considerando que sus resultados no eran representativos ni comparables. En el Anexo II se presenta un detalle de los problemas en la aplicación.

³³ Continúa el trabajo iniciado en el marco del proyecto MAPEAL (mapeal.cippec.org), que buscó comprender los resultados y las políticas educativas de siete países en América Latina entre 2000 y 2015, plasmado en el libro *América Latina después de PISA*.

Gráfico 24. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012



(*) Se presentan los resultados de 2015 a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral. Por lo tanto, sus resultados no se consideran comparables.

Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2012.

Los resultados de las pruebas en el país perfilan en general una tendencia estable si se considera la comparación puntual 2000 - 2012, con una leve reducción (pasa de 401 puntos a 397 en ambos años, en el promedio de las tres áreas), pero se identifican variaciones significativas al interior del período. Principalmente, se resalta la caída en los desempeños entre 2000 y 2006 de 19 puntos promedio, que es seguida de una recuperación 2006 - 2009 de 14 puntos promedio.

En la última aplicación válida de Argentina, el país se encuentra por debajo del promedio de la región, por debajo de Chile, México y Uruguay, y con valores similares a Brasil y Colombia. En un marco donde los países de la región han mejorado sus desempeños, esta estabilidad deviene en una pérdida relativa de posiciones: al inicio del ciclo PISA, Argentina manifestaba los desempeños más altos de la región junto con Chile y México.

Los resultados reescalados ponen en duda este perfil, poniendo de manifiesto que la evolución 2006 - 2012 de Argentina no muestra una tendencia de mejora sostenida. Por el contrario, considerando los resultados promedio de las tres áreas con la metodología 2015, Argentina obtuvo en 2012 desempeños similares al 2006, e inferiores al 2009. Lamentablemente no es posible analizar qué resultado hubiese obtenido el país en el 2000 con esta nueva metodología.

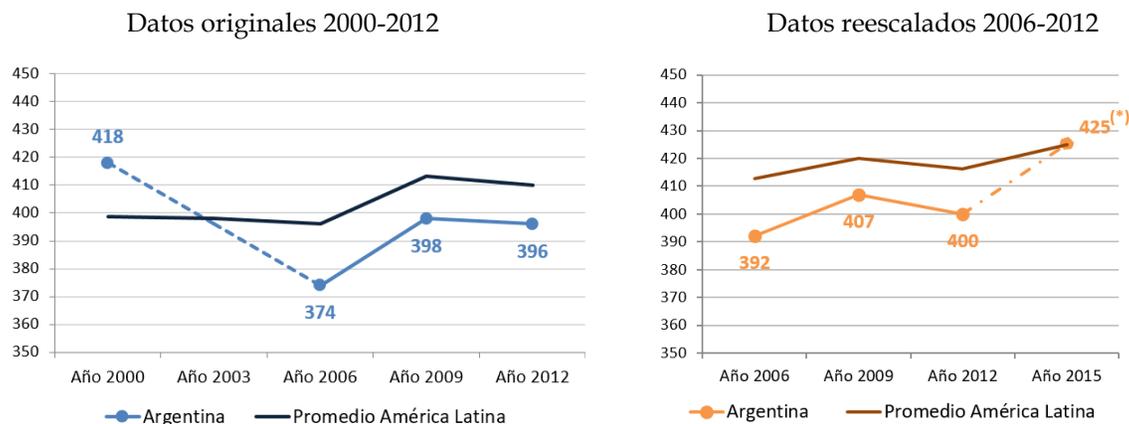
Los resultados del 2015 - no comparables directamente con los años anteriores - implicarían una mejora sustantiva de los desempeños respecto del 2012, y posicionarían al país por encima del promedio regional. En el anexo II se profundiza en la caracterización de estos resultados y en hipótesis para su interpretación.

Argentina se caracteriza por ser uno de los países de la región con mayor presencia de evaluaciones incompletas. Tal como fue analizado en el capítulo 2, los porcentajes de *non-reached*

ítems son altos en las sucesivas aplicaciones de PISA. Tanto en PISA 2000 como en 2006 y 2009, más del 10% de los ítems evaluados pertenecen a esta categoría.

Esta tendencia resumen de desempeños se complejiza al abordar la caracterización de cada área, ya que asumen comportamientos muy diferentes.

Gráfico 25. Puntaje promedio en lectura, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012

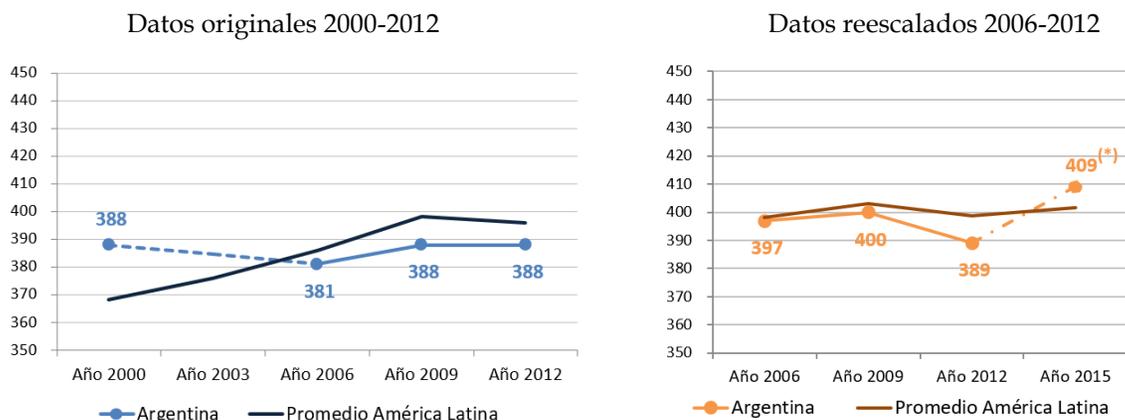


(*) Se presentan los resultados de 2015 a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2012.

En el caso de lectura, los resultados iniciales del 2000 eran muy alentadores, pero en el ciclo 2006 Argentina empeoró marcadamente sus desempeños. Observando los resultados reescalados, la mejora 2006 - 2012 es de 8 puntos, diferencia que no es estadísticamente significativa (con los resultados originales, se identificaba una mejora significativa de 22 puntos). El escenario que se perfila es de cierta estabilidad.

Gráfico 26. Puntaje promedio en matemática, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012

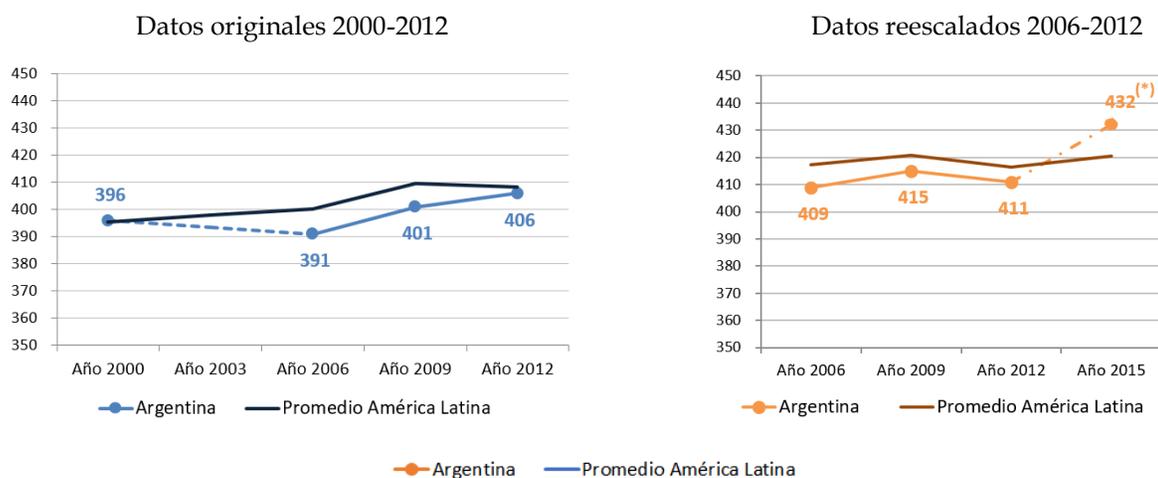


(*) Se presentan los resultados de 2015 a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2012.

En matemática, los resultados originales del 2000 mostraban niveles muy similares a los que obtuvo el país en cada una de las evaluaciones en las que participó. Los resultados reescalados muestran que la tendencia de Argentina ha sido diferente: en el año 2012 Argentina obtiene 389 puntos, los resultados más bajos de las últimas tres aplicaciones. La caída 2009 - 2012 de 11 puntos es un retroceso estadísticamente significativo (que en los resultados originales perfilaba estabilidad).

Gráfico 27. Puntaje promedio en ciencias, resultados originales y reescalados. Argentina y promedio de países seleccionados de América Latina. 2000-2012



(*) Se presentan los resultados de 2015 a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2012.

En el área de ciencias, los resultados originales del 2000 se posicionan en valores bajos, y en los años posteriores el país mejora sus desempeños. Ya con los resultados calculados bajo la metodología de 2015, la tendencia 2006 – 2012 es de marcada estabilidad, se observa una diferencia positiva de apenas 2 puntos en el periodo. La mejora que se habría observado con los resultados originales, de 14 puntos, no era estadísticamente significativa.

Las evaluaciones nacionales de Argentina permiten algunas comparaciones con estas tendencias. Identificar similitudes o diferencias ofrece información para contrastar el diagnóstico.

Recuadro 3. Las evaluaciones nacionales de la calidad educativa en Argentina

La política de evaluación en el país ha sufrido varias modificaciones a lo largo de su existencia, que inciden en la comparabilidad de los resultados, y generan dificultades para la construcción de series de tiempo.

El país implementó el Operativo Nacional de Evaluación (ONE) en el año 1993, el cual se aplicó en forma anual hasta el año 2000, bianual hasta 2007, y trianual hasta 2013. En el año 2016 fue reemplazado por la evaluación Aprender, de aplicación anual.

Desde 1996, las evaluaciones ONE se aplican regularmente al 3er y 6to grado de la primaria, y en el 9vo y 12vo año de estudio (actualmente equivalentes al 2do/3er año de la secundaria y al 5to/6to año³⁴). La evaluación ONE fue muestral en la mayoría de las aplicaciones, en algunas de ellas se optó por implementarla en forma censal (para el 12 año en 1997, 1998, 1999, 2000, 2010, 2013; y para 6to año de primaria en el 2000). Aprender implementó en forma censal la evaluación para estos dos años de estudio (6to y 12vo) y muestral en las otras dos áreas. La aplicación 2017 sólo será censal en el 12vo año.

ONE evaluó cuatro áreas para todos los años de estudio: lengua, matemática, ciencias Naturales y ciencias Sociales. Aprender 2016 mantuvo estas cuatro áreas sólo en el 12vo año, y redujo a lengua y matemática las áreas evaluadas en los otros años de estudio. En ambos casos, las pruebas están basadas en el currículum (los Núcleos de Aprendizaje Prioritario).

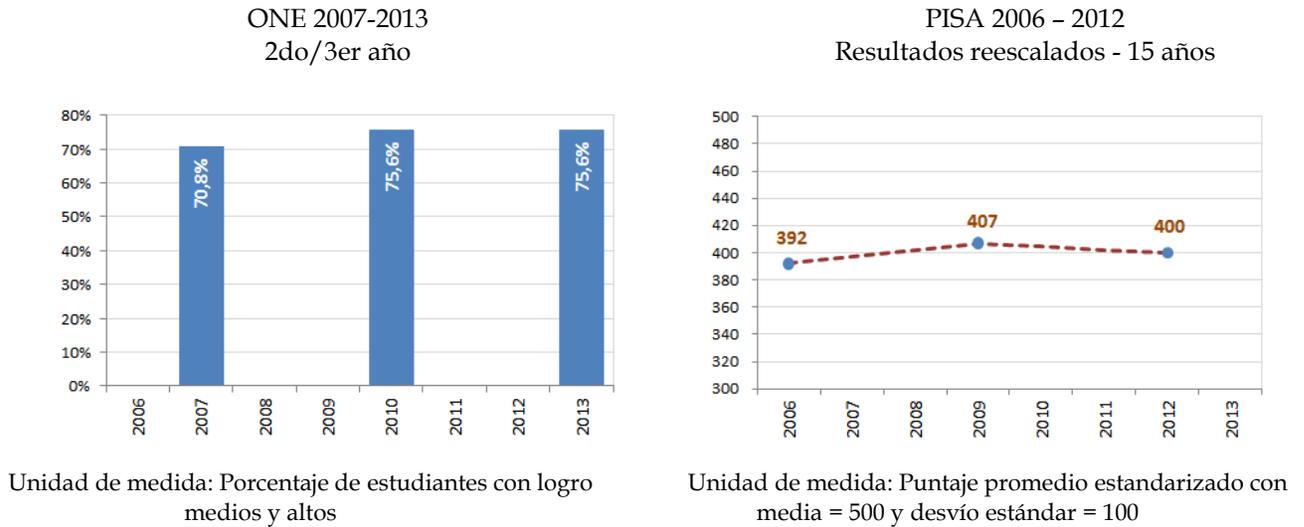
En la aplicación del año 2005 el diseño del ONE se modifica, implementando un modelo de prueba criterial basada en la Teoría del Respuesta al Ítem, lo que impide la comparabilidad con los años anteriores. Sus resultados se expresan en tres categorías de logro: Porcentaje de estudiantes con logros básicos, satisfactorios y avanzados. Aprender mantiene el diseño general de las pruebas, pero modifica las categorías de logro, ampliando a cuatro opciones: por debajo del nivel básico, nivel básico, nivel satisfactorio y nivel avanzado.

La comparabilidad de los resultados se ve afectada por otros factores: Los resultados 2005 – 2007 no son directamente comparables. Desde ONE 2007 a ONE 2013 los resultados se consideran comparables, aunque no hay publicaciones oficiales que lo establezcan. Aprender 2016 incluyó en su publicación el recalcule de los resultados ONE 2013 bajo las nuevas cuatro categorías, pero este criterio no se aplicó en años anteriores. Por lo tanto, el país dispone de una serie de resultados 2007 – 2013 basados en las pruebas ONE, y otra serie 2013 – 2016 basada en los resultados de Aprender.

34 En Argentina, desde la Ley de Educación Nacional 26.206(año 2006) coexisten dos estructuras de niveles: algunas jurisdicciones se organizan con 6 años de educación primaria y 6 de secundaria, y otras con un año más de primaria y uno menos de secundaria (7 y 5 respectivamente). Las jurisdicciones se distribuyen en forma igual entre ambas, lo que obliga a mantener una nomenclatura equivalente para ambas estructuras.

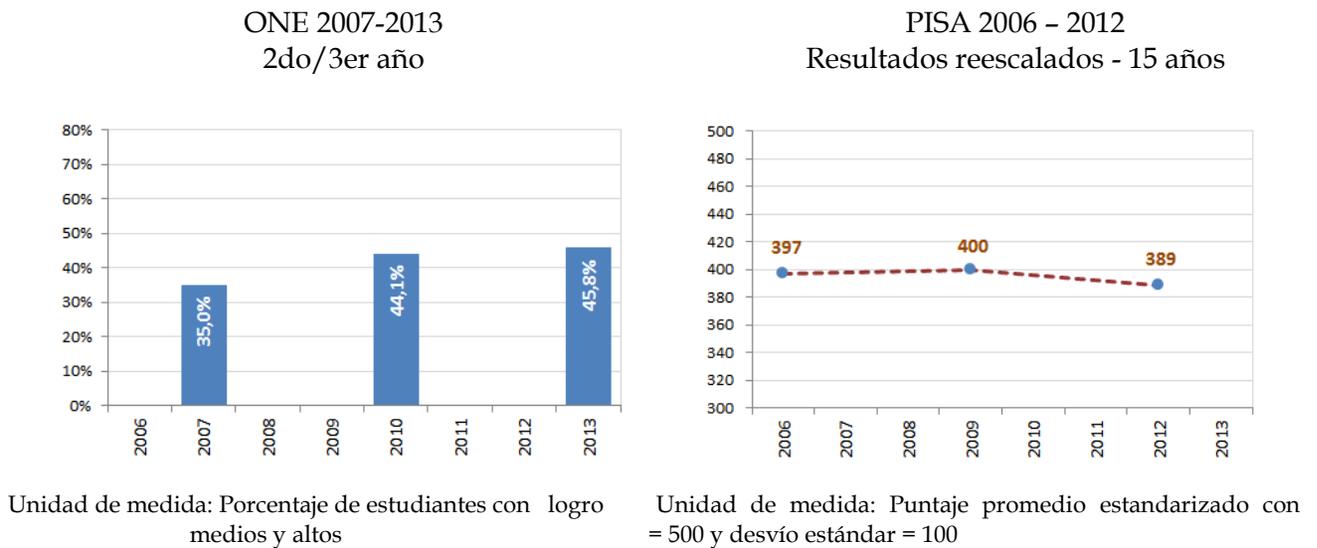
Los resultados de las pruebas ONE del 2do/3er año de la secundaria pueden ser utilizados como información para contrastar las tendencias que presenta PISA, ya que la edad teórica para el grado son los 14 años de edad, población cercana a la que evalúa la OCDE.

Gráfico 28. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Argentina. 2006-2013



Fuente: CIPPEC, sobre la base de resultados ONE 2007, 2010 y 2013 y OCDE-PISA 2006-2012.

Gráfico 29. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Argentina. 2006-2013



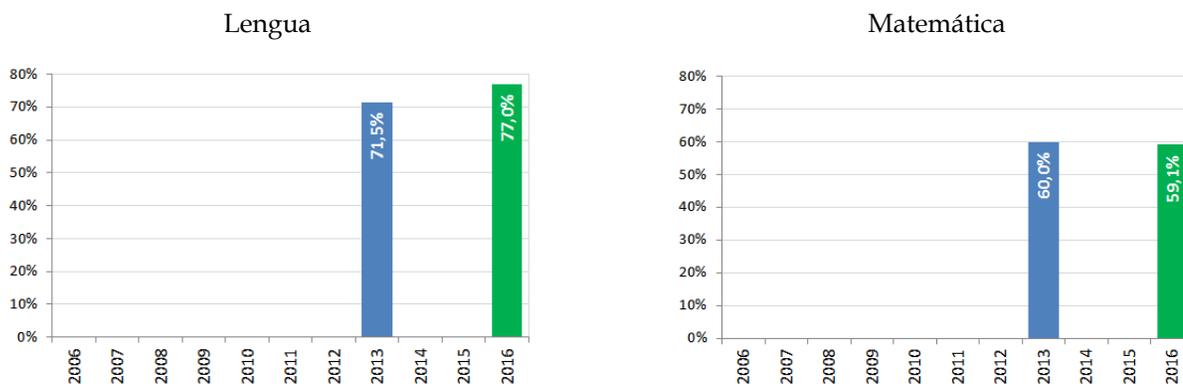
Fuente: CIPPEC, sobre la base de resultados ONE 2007, 2010 y 2013 y OCDE-PISA 2006-2012.

En el área de lectura, las pruebas ONE presentan una tendencia levemente creciente entre 2007 y 2010, evidenciándose en un incremento del porcentaje de estudiantes con logros medios y altos del 70,8% al 75,6%. En el período siguiente la evaluación muestra un estancamiento, con resultados similares entre 2013 y 2010.

En matemática el perfil es similar, aunque el punto de partida es mucho más bajo: en el ONE 2007 solamente el 35% de los estudiantes manifestaba logros medios y altos. Este valor mejoró significativamente en el 2010, alcanzando al 44,1% del total de evaluados. Para el 2013 se reiteran valores muy similares, con un muy escaso incremento de 1,7 puntos porcentuales.

Tal como se ha mencionado en el **recuadro 3** con la aplicación de las evaluaciones Aprender implementadas en el 2016, el país ha calculado una medida comparable con los ONE para identificar cambios en los desempeños entre 2013 y 2016. Esta comparación se elaboró sobre las pruebas aplicadas al 5to/6to año (12vo año de estudios), que es una población menos comparable con PISA: por la edad teórica del grado (17 años), por el hecho de que la acumulación de sobreedad incluye en el grado a numerosos estudiantes de 18 y 19 años, y por el significativo porcentaje de estudiantes fuera de la escuela en esta edad.

Gráfico 30. Resultados ONE 2013 - Aprender 2016. 5to/6to año. Argentina. 2013-2016



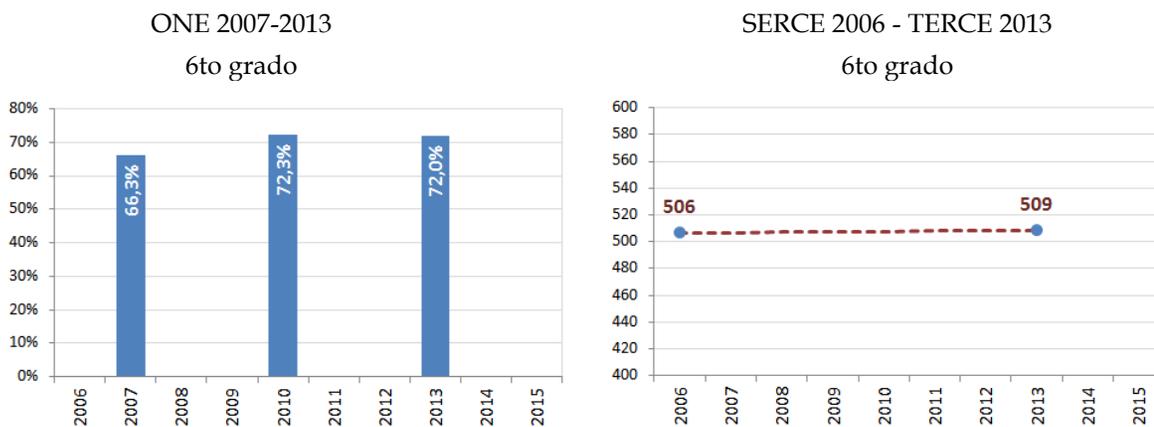
Unidad de medida : Porcentaje de estudiantes con logro básicos, satisfactorios y avanzados.

Nota: las aplicaciones ONE 2007 y 2010 no se publicaron con el mismo escalamiento de Aprender 2016.

Fuente: CIPPEC, sobre la base de resultados ONE 2013 y Aprender 2016

Teniendo en cuenta estas advertencias, es posible identificar en la comparación una mejora en los desempeños en lectura de los estudiantes entre 2013 y 2016. Tomando la unidad de medida de la evaluación Aprender (porcentaje de estudiantes con logro básicos, satisfactorios y avanzados) se identifica que los resultados del país se incrementaron del 71,5% al 77,0% en el área. Sin embargo, en matemática no se identificaron mejoras. Por el contrario, los resultados disminuyen levemente del 60,0% al 59,1% en el período.

Gráfico 31. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2013

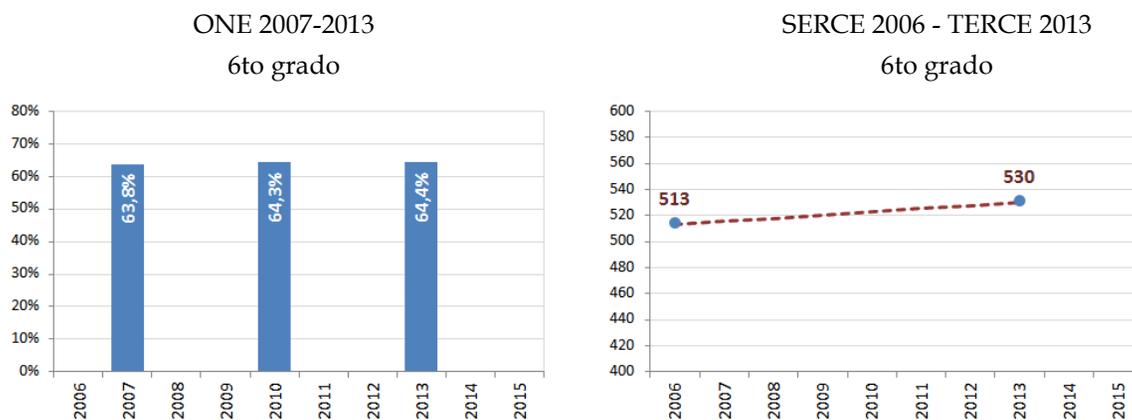


Unidad de medida: Porcentaje de estudiantes con logro medios y altos

Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados ONE 2007, 2010 y 2013 y SERCE - TERCE.

Gráfico 32. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2013



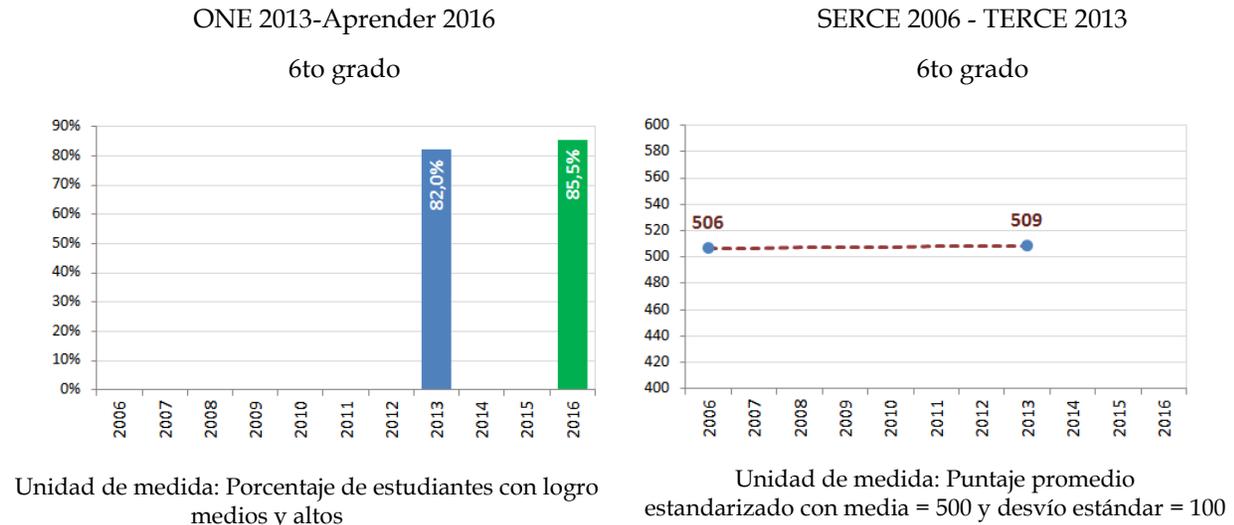
Unidad de medida: Porcentaje de estudiantes con logro medios y altos

Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados ONE 2007, 2010 y 2013 y SERCE - TERCE.

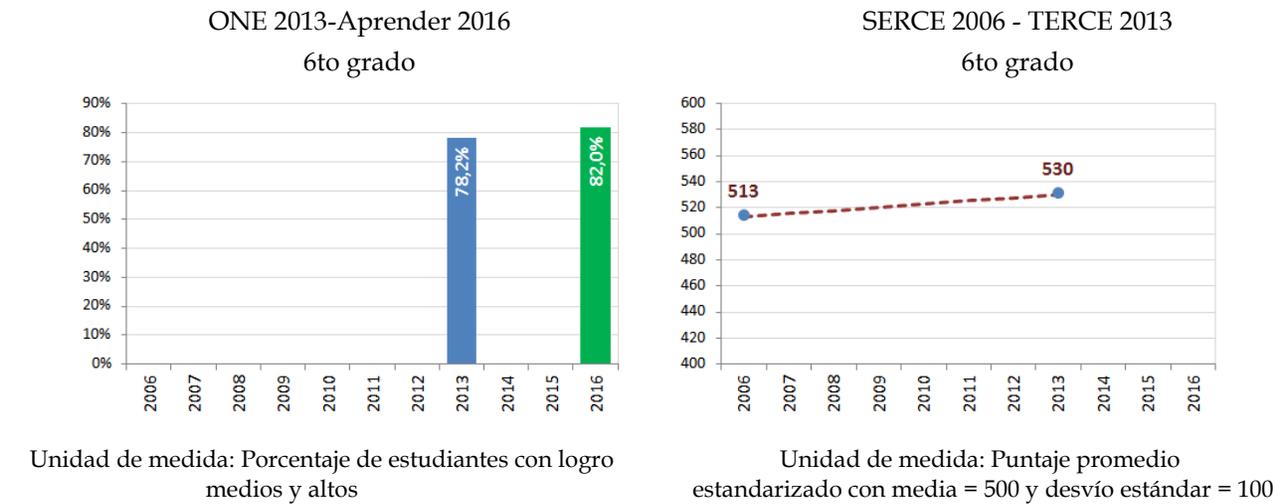
Otros resultados interesantes para incorporar al escenario de análisis de los desempeños de los estudiantes son los del nivel primario. Los resultados del ONE 2007 - 2013 para el 6to grado de la primaria muestran tendencias similares a las ya analizadas en secundaria: se observa una mejora en los resultados en lengua entre 2007 y 2010 (incremento del 66,3% al 72,3% de resultados medios y altos), pero seguidos de un estancamiento hacia 2013 (72,0%). En matemática los resultados se han mantenido marcadamente estables, con un incremento de 0,6 puntos porcentuales entre 2007 y 2013.

Gráfico 33. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2016



Fuente: CIPPEC, sobre la base de resultados ONE 2013, Aprender 2016 y SERCE - TERCE.

Gráfico 34. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Argentina. 2006-2016



Fuente: CIPPEC, sobre la base de resultados ONE 2013, Aprender 2016 y SERCE - TERCE.

Los resultados de Aprender, en cambio, muestran un repunte en ambas áreas. Tanto en lengua como en matemática, los resultados del 2016 son superiores al 2013. Medidos en porcentaje de estudiantes con logros básicos, satisfactorios y avanzados, los resultados del grado se incrementaron del 82,0% al 85,5% en lengua, y del 78,2% al 82,0% en matemática.

En el 6to grado de la primaria también se aplicaron las pruebas TERCE del LLECE-UNESCO, lo que permite contrastar estas tendencias con una prueba internacional. Los resultados del 2013 posicionan al país en lugares intermedios en la región, por encima del promedio de países. En el escenario internacional los resultados han sido mejores en matemática. En cambio, en lectura se sitúa por debajo de la mayoría de los 7 países seleccionados y analizados en este documento. La mirada comparada SERCE-TERCE permite identificar cambios en los desempeños por área: el país incrementó sus puntajes en matemática entre 2006 y 2013 en magnitudes similares al conjunto de países: pasó de 513 a 530 puntos (+17 puntos), levemente inferior a la mejora de 19 puntos en el conjunto de los 15 países participantes. En lectura el país no evidencia cambios significativos: los puntajes apenas se incrementaron entre ambas aplicaciones, de 506 a 509 puntos.

Los escenarios que construyen las evaluaciones nacionales e internacionales tienden a ciertas conclusiones convergentes: en general, los resultados muestran una situación de relativa estabilidad, desempeños que no han mejorado ni empeorado significativamente al menos entre la segunda mitad de la década del 2000 y los primeros años de la siguiente. Hay coincidencia entre evaluaciones en subrayar el hecho de que en el nivel secundario las dificultades parecen concentrarse en matemática; en cambio, en primaria los diagnósticos entre las pruebas nacionales e internacionales no son confluyentes.

Los resultados de Argentina 2015 constituyen una incógnita: tal como se explica en el anexo II, no hay evidencias claras de que los problemas en la muestra hayan sesgado fuertemente los resultados. Bajo esta hipótesis, se podría inferir que Argentina podría haber obtenido resultados similares – y representativos – si no hubiese cometido errores técnicos en la confección de la muestra. Por su parte, la evaluación Aprender de 2016 muestra una mejora generalizada de los desempeños respecto del 2013, lo que sería relativamente consistente a los resultados de PISA 2015, aunque con un incremento de menor magnitud. Sin embargo, los cambios implementados en la política de evaluación nacional, junto con la mejora en la cobertura del operativo (Secretaría de Evaluación Educativa, 2016), son factores que también podrían influir en esta comparación. En síntesis, existen indicios de mejora en los desempeños de los estudiantes argentinos en los últimos años, aunque no es posible dimensionarlos en forma precisa debido a aspectos metodológicos que dificultan la comparación.

Resumen de datos educativos y sociales destacados de Argentina

Cobertura

1. Argentina se destaca como un país con una historia de acceso educativo ya consolidada desde al menos cuatro décadas. Las tasas de analfabetismo eran las segundas más bajas del conjunto de países analizados, luego de Uruguay, ya en 1980 (6,1%). El país alcanza tempranamente altas tasas de cobertura en el nivel primario: 94,8% en 1970, 99,4% en 2010.
2. En el nivel secundario muestra también altas tasas netas de cobertura, con un 86,6% en el 2014, los valores más altos en el conjunto de países analizados. El crecimiento en la cobertura

educativa en el nivel secundario durante el período analizado fue intermedio, partiendo de un nivel ya alto para la región.

3. Argentina también se destaca por sus avances de la cobertura en el nivel inicial, aunque en valores ya similares a la mayoría de los países analizados: entre 2000 y 2014 la tasa bruta de matrícula de 3 a 5 años se incrementa del 44% al 75% (SEDLAC).

Trayectorias

1. Las trayectorias de los alumnos se ubican en un punto intermedio entre los países estudiados, con relativamente altos niveles de repitencia y sobreedad.
2. En el nivel primario, los niveles de repitencia en el año 2000 se situaban levemente por debajo del promedio de países (6,1% de repitentes en Argentina, y 8,6% en el conjunto de los 7 países). Al año 2015 la repitencia se redujo drásticamente en el nivel, alcanzando a apenas el 2,0% de la matrícula. Esta reducción también fue identificada en la región, donde el guarismo descendió al 3,7%.
3. En el ciclo básico de la secundaria, los valores al inicio del período también situaban al país cerca del promedio (8,1% en Argentina y 8,4% en el conjunto). En ambos casos, y contrariamente a lo que ocurrió en primaria, la repitencia tiende a subir. En este nivel, el incremento de la repitencia en Argentina (que alcanza al 11,0% de la matrícula) fue superior al conjunto de países (que llegó al 9,1%)
4. La trayectoria de los alumnos atravesó una situación similar a la de la cobertura. Históricamente hubo grandes niveles de repetición: al año 2001, el 25% de los estudiantes de primaria eran repetidores, valores muy altos en relación al promedio de los 7 países seleccionados (8,5%) (UIS). Sin embargo, para el año 2011 hubo una mejora los niveles de repitencia alcanzando el 8,7% (UIS), aunque muy por encima del promedio de los 7 países (3,6%) (UIS- circa 2015).

Equidad

1. Argentina muestra en los resultados del año 2012 niveles bajos de desigualdad en PISA, con 81 puntos de diferencia en los resultados entre el cuartil I y el IV (promedio lectura y matemática). En lo que se destaca el país es que ha sido el que más ha reducido la brecha de desigualdad entre el 2000 y el 2012, con una disminución promedio de 31 puntos en un contexto donde la región incrementa la brecha en 18 puntos.
2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos de 2009 (Vázquez, 2012) muestra que Argentina ocupa lugares intermedios en sus niveles de segregación educativa vinculada al nivel socioeconómico, dentro del universo de los siete países estudiados. A su vez, es el segundo país que más ha reducido los niveles de segregación tomando como punto de comparación el año 2000.

Contexto

1. Durante el siglo XXI se perfila una mejora significativa en las condiciones de vida de la población, concentrada en el segundo quinquenio. Según las estadísticas oficiales del país, el porcentaje de población bajo la línea de pobreza ascendió del 28,9% en 2000 al 31,4% en 2006, y luego se incrementó nuevamente al 32,2% en 2015, aunque existen ciertas dudas sobre la

comparabilidad de los datos. Un indicador comparable en la región, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, muestra una mejora en el periodo, del 39,3% en el 2000, al 24,6% en 2015. A su vez, el indicador de pobreza construido por PISA (índice del status económico, social y cultural, ESCS por sus siglas en inglés) perfila también una mejora en las condiciones de vida de la población de 15 años entre 2000 y 2009, con una estabilidad posterior.

2. Esta evolución de la pobreza es consistente con el comportamiento de los indicadores macroeconómicos durante la década del 2000: partiendo de valores muy bajos en los primeros años del 2000, el PBI de la Argentina creció sostenidamente desde el 2003 hasta el 2011, en valores constantes (IMF-WEO), con incrementos anuales cercanos al 8%, a excepción de los años 2008 (3% de incremento) y 2009 (sin incremento del PBI). A partir del año 2012 se manifiesta una desaceleración intensa que se mantiene hasta la actualidad. Entre 2011 y 2015 el PBI, a valores constantes, apenas un 5,4%.
3. En estos años de crecimiento económico ha mejorado en forma marcada la distribución social de la riqueza, medida a través del coeficiente de Gini, con un importante descenso del 0,50 al 0,42 (es el país con mayores mejoras en la distribución del ingreso entre 2000 y 2014). Al año 2014, Argentina tenía, junto con Uruguay, los niveles más bajos de desigualdad en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados es de 0,47 (SEDLAC).
4. Al año 2014, Argentina también mostraba niveles intermedios de empleo formal (el 33% de los trabajadores no poseen cobertura de pensión, contra el 34% promedio de los siete países), niveles intermedios de desempleo en 2015 (6,5% en el país y 6,7% en el promedio), y niveles intermedios de cobertura de salud (el 73,2% de la población ocupada urbana posee cobertura de salud, levemente superior al promedio de los 7 países, de 70,4%)

Inversión educativa

1. Argentina aumentó marcadamente su inversión en el sistema educativo en los últimos años. Medido como porcentaje del PBI, el gasto en educación había disminuido del 4,6% en el 2000 al 3,2% en 2004, y luego mejoró significativamente al 6,1% en 2015³⁵. Considerando la comparación entre años inicial y final, este crecimiento es de magnitud intermedia en América Latina en el periodo estudiado. Al año 2012 alcanza niveles de inversión (medidos como % del PBI) altos en la región (el promedio de 16 países de la región es 4,6% *circa* 2012).
2. Asimismo, el país destina, dentro del gasto público total, el 11,6% a educación (año 2013). En el promedio de 16 países de la región este porcentaje era del 14,3% en valores *circa* 2012. Esta porción de los recursos públicos destinados a educación mejora levemente en relación al 2002, año en el que se destinó a educación el 10,3% del gasto público (no se consideran los años 2000 y 2001, que asumen valores excepcionales a causa de la crisis).

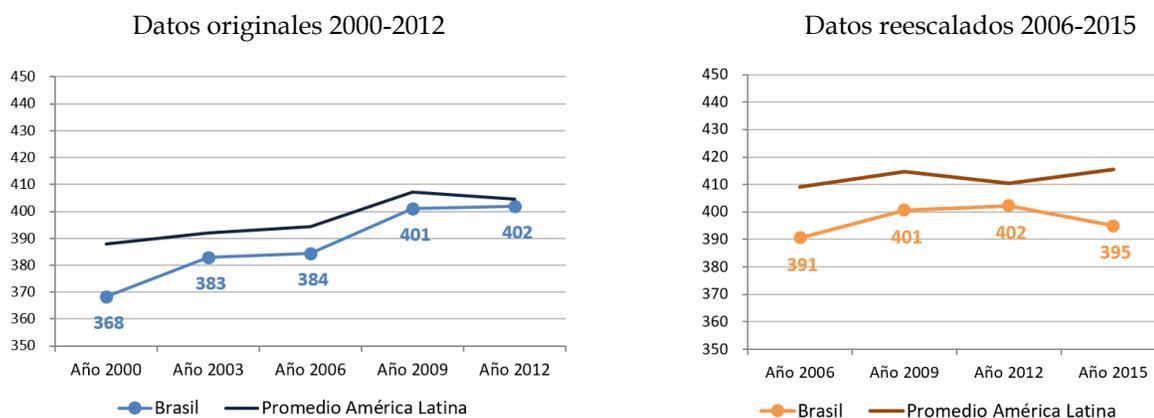
³⁵ Fuente de datos: CIPPEC. Los cálculos de 2015 fueron realizados sobre la base del cálculo del PBI realizado en 2004.

4.2. Brasil

Resultados comparados de las evaluaciones de la calidad educativa

En las pruebas PISA, Brasil participó en forma continua desde el año 2000 al 2015 en las sucesivas aplicaciones. Los resultados de las pruebas en el país perfilan en general una leve mejora durante el periodo 2000 – 2009 y un posterior estancamiento. Si bien en la última aplicación se encuentra junto con Perú en las escalas más bajas de resultados de la región, ha logrado reducir las brechas que lo separan de los otros países analizados, en comparación al inicio del período.

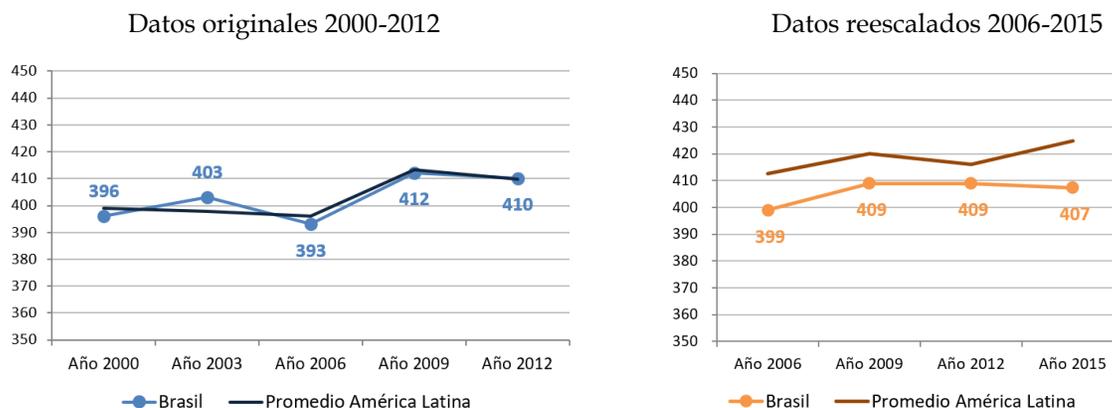
Gráfico 35. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

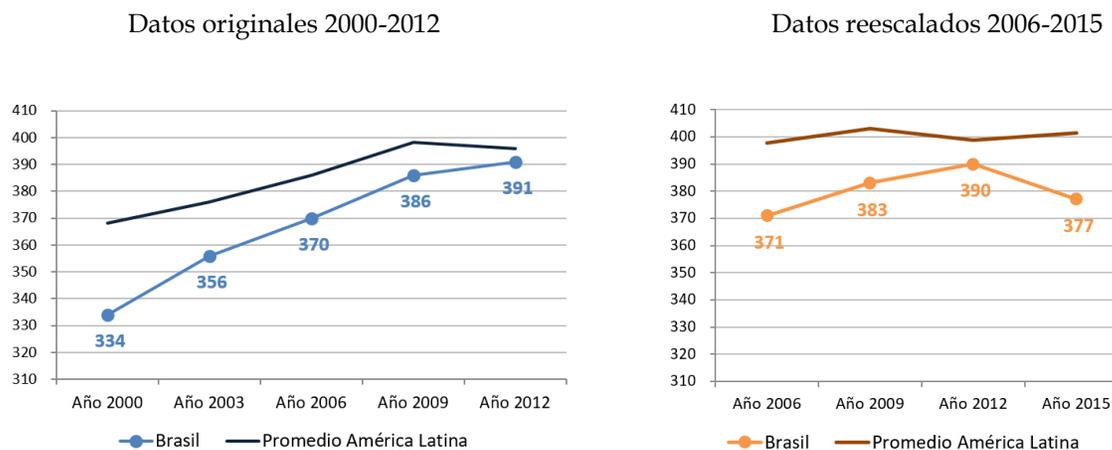
La mejora entre 2000 y 2009 es en promedio de 33 puntos, diferencia superior al promedio de países evaluados en este tramo (25 puntos). Estas diferencias son estadísticamente significativas en las tres áreas evaluadas. Sin embargo, en las últimas aplicaciones de PISA los resultados de Brasil no se han incrementado, e incluso muestran una tendencia decreciente. En promedio se observa un retroceso de 6 puntos, que asume su punto más crítico en el área de matemática, donde el país disminuye sus desempeños de 389 a 377 puntos, diferencia estadísticamente significativa. Es en esta área donde Brasil obtiene en 2015 los resultados más bajos en el conjunto de los 7 países analizados.

Gráfico 36. Puntaje promedio en lectura, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Gráfico 37. Puntaje promedio en matemática, resultados originales y reescalados. Brasil y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

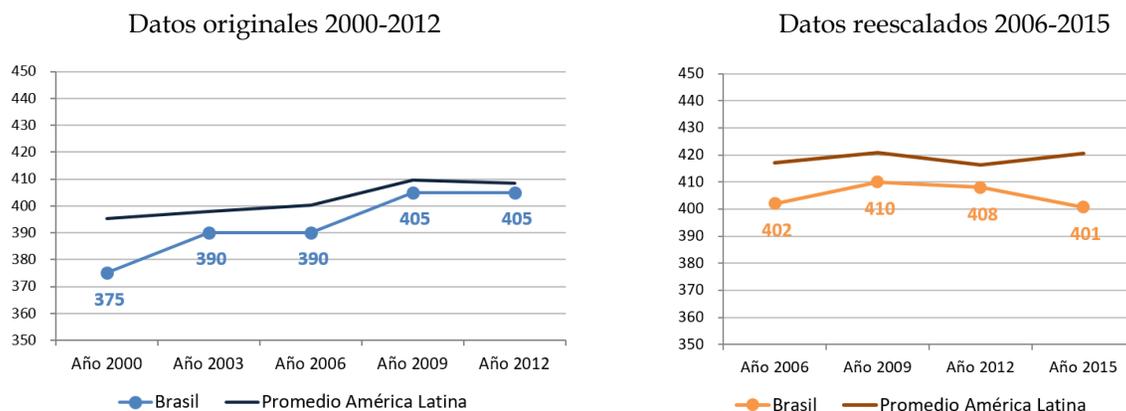
Este estancamiento evidenciado en los últimos años se corrobora al considerar los resultados reescalados con el cambio metodológico implementado en el último ciclo de PISA, disponibles desde el 2006 en adelante (tal como se presenta en el capítulo 2). Tanto en matemática como en lectura, los resultados reescalados del período muestran, con algunas leves diferencias, las mismas tendencias.

Como ocurre con los demás países, no es posible estimar cuánto impactaría este cambio de metodología en la comparación de resultados PISA del año 2000. Es posible suponer que los resultados reescalados previos al 2006 podrían mostrar diferencias marcadas con los publicados. Como fue presentado en el capítulo 2, Brasil manifestaba altos porcentajes de *non-reached items* en PISA 2000 y 2003.

Otro aspecto a tener en cuenta en la comparación en el tiempo de los resultados de PISA es la observación realizada por Klein (2011). El autor demuestra que la definición de fecha de referencia

para establecer la población objetivo de PISA fue variando en el tiempo en Brasil: en el 2000 se consideraron los años cumplidos al 30 de junio. En 2003 y 2006, al 30 de abril. En 2009, al 31 de diciembre. El autor muestra que estos cambios impactaron en la distribución de la muestra por grado. Si se corrige el sesgo provocado por este cambio en la distribución, las mejoras observadas en el país en las tres áreas entre 2006 y 2009 (17 puntos en promedio) casi desaparecen (sólo 1 punto en promedio).

Gráfico 38. Puntaje promedio en ciencias, resultados originales y reescalados. Nivel primario y secundario. Brasil y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

En el caso de ciencias, el uso de la nueva metodología incide marcadamente en los análisis que pueden hacerse sobre los resultados. La OCDE, en sus notas destacadas del país, señala que los resultados en el área se han mantenido estables desde 2006. El incremento de 10 puntos, desde 390 a 401 en el periodo, no es estadísticamente significativo, y es similar al promedio de países OCDE (OCDE, 2016c). Pero este incremento se reduce con los reescalamientos de resultados que PISA realiza a partir de los cambios metodológicos implementados en 2015: la diferencia de puntajes se reduciría a sólo 1 punto si se calcula el resultado del 2006 con la nueva metodología.

Otro aspecto resaltado es el incremento de la tasa de escolarización a los 15 años de edad, que en 2015 es del 71%, 15 puntos superior al 2003. Esta información se destaca en el informe nacional de PISA 2015, y permite remarcar el logro de haber incrementado la cobertura sin una caída significativa en los resultados de PISA (INEP, 2016).

Las evaluaciones nacionales de Brasil permiten establecer una instancia de comparación con estos diagnósticos. El **recuadro 4** describe el sistema nacional de evaluación de la calidad educativa del país.

Recuadro 4. Las evaluaciones de la calidad educativa en Brasil

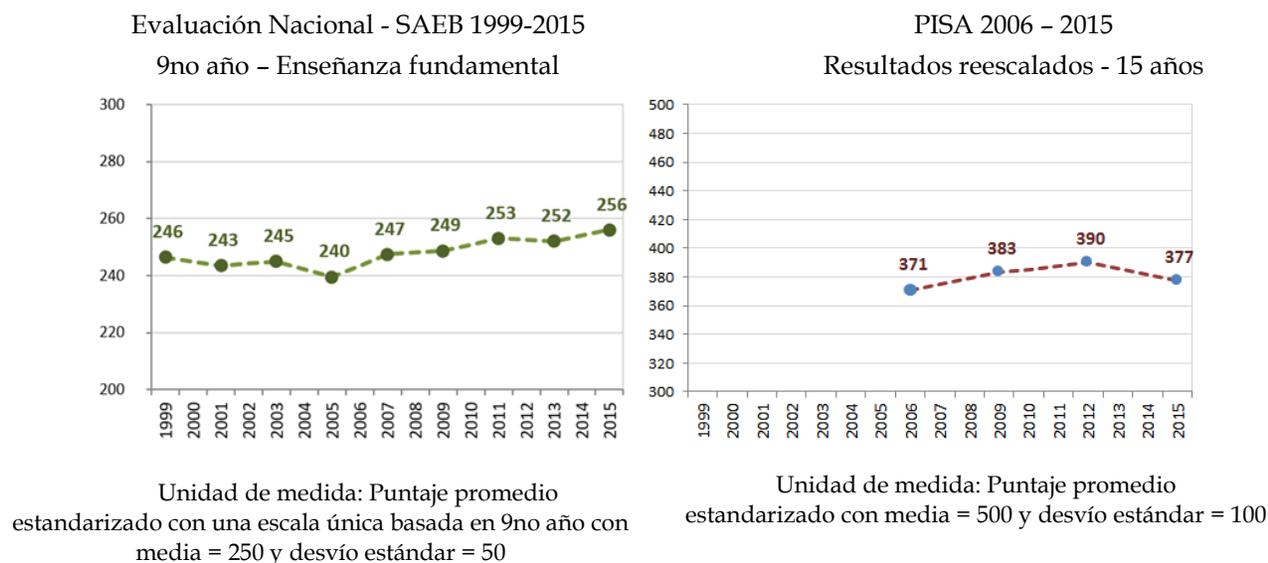
Brasil desarrolló el Sistema de Evaluación de la Educación Básica (SAEB), que actualmente se compone de tres evaluaciones de gran escala:

- Evaluación Nacional de la Educación Básica (ANEB), una evaluación muestral de estudiantes de 5° y 9° año de la enseñanza fundamental, y en el 3° año de la enseñanza media, que asisten a escuelas públicas y privadas;
- Evaluación Nacional del Rendimiento Escolar (ANRESC), evaluación censal de estudiantes del 5° y 9° años de la enseñanza fundamental de las escuelas públicas de las redes municipales, estatales y federal;
- Evaluación Nacional de la Alfabetización (ANA), una evaluación censal de los estudiantes del 3° año de la Enseñanza Fundamental que asisten a escuelas públicas.

Tanto la ANEB como la ANRESC evalúan lectura y matemática, mientras que la prueba ANA se enfoca en alfabetización en lengua portuguesa y matemática. Las evaluaciones son diseñadas para evaluar habilidades, están alineadas con el currículum y son referidas a criterios. Comparten una misma escala de puntajes, que se define en base al 9° año de la enseñanza fundamental con una media de 250 puntos y desvío estándar de 50 puntos.

Los resultados de las pruebas SAEB de 9no año de la enseñanza fundamental pueden ser comparados con las tendencias de PISA en términos de desempeños en lectura y matemática. La edad teórica para este grado son los 14 años de edad, lo que establece cierta equivalencia de comparación con la población objetivo de PISA, considerando los altos niveles de sobreedad en el país.

Gráfico 39. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015

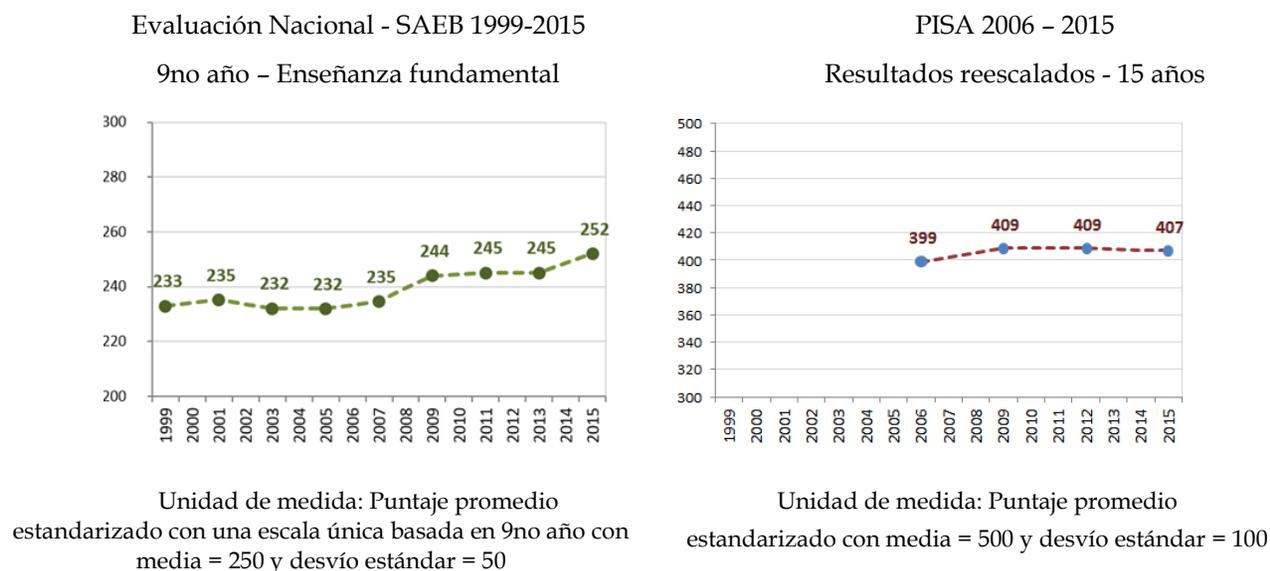


Fuente: CIPPEC, sobre la base de resultados SAEB 1999-2015 y OCDE-PISA 2006-2015.

Para este grado, los resultados del SAEB muestran una tendencia relativamente estable en la primera mitad de la década del 2000 tanto en lectura como en matemática, en este segundo caso con cierta caída en los puntajes (de 246 en 1999 a 240 en 2005).

En ambas áreas, este estancamiento inicial se revierte en la segunda mitad de la década: desde el año 2005 en adelante los desempeños han ido incrementándose en forma continua. En lectura, los resultados aumentan 20 puntos entre 2005 y 2015 (de 232 a 252 puntos); y en matemática el incremento es de 16 puntos en el mismo periodo (240 a 256 puntos). Ambas diferencias son importantes, considerando que el desvío estándar de la escala es de 50 puntos.

Gráfico 40. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario Brasil. 1999-2015



Fuente: CIPPEC, sobre la base de resultados SAEB 1999-2015 y OCDE-PISA 2006-2015.

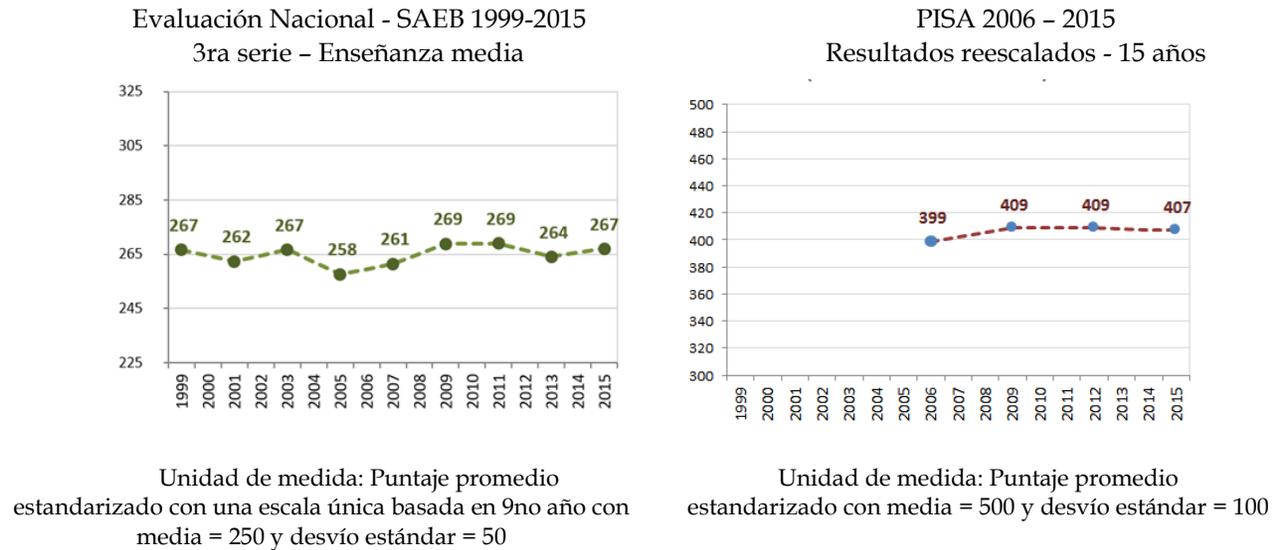
En lectura, los resultados de la evaluación nacional de 9º muestran una tendencia a la mejora entre 2001 y 2010, periodo en el que los puntajes se incrementan de 252 a 259 puntos. En los años posteriores, hay un marcado cambio de sentido: los resultados de lectura tienden a una caída sostenida, alcanzando en el 2015 y 2016 los valores más bajos de las últimas dos décadas, con 247 puntos. En matemática, la tendencia es más promisoriosa: se observa una mejora sostenida en el período 2001 - 2016, con una diferencia neta total de 18 puntos. Se visualiza un crecimiento prácticamente constante desde el año 2001 (248 puntos) hasta el año 2013 (267 puntos). En los años posteriores, el crecimiento se frena, llegando a los 266 puntos en 2016.

El informe nacional de PISA 2015 de Brasil incluye un capítulo donde se realiza un análisis comparado de las pruebas PISA y de SAEB de 9º año para el año 2015. La conclusión principal a la que se arriba es que las evaluaciones presentan importantes diferencias, no sólo en las matrices de referencias, sino también en la estructura y metodologías empleadas (INEP, 2016), lo que dificulta la comparación directa. Sin embargo, a pesar de tener focos distintos, ambas evaluaciones tienen ciertos planos de confluencia en el retrato que construyen de la educación básica brasileña en 2015. El

informe muestra que los niveles de correlación entre PISA 2015 y la prueba SAEB (en este caso ya para el 3° año de la enseñanza media) son muy altos, y confluyen en la caracterización de la diferencia de resultados entre los diferentes territorios.

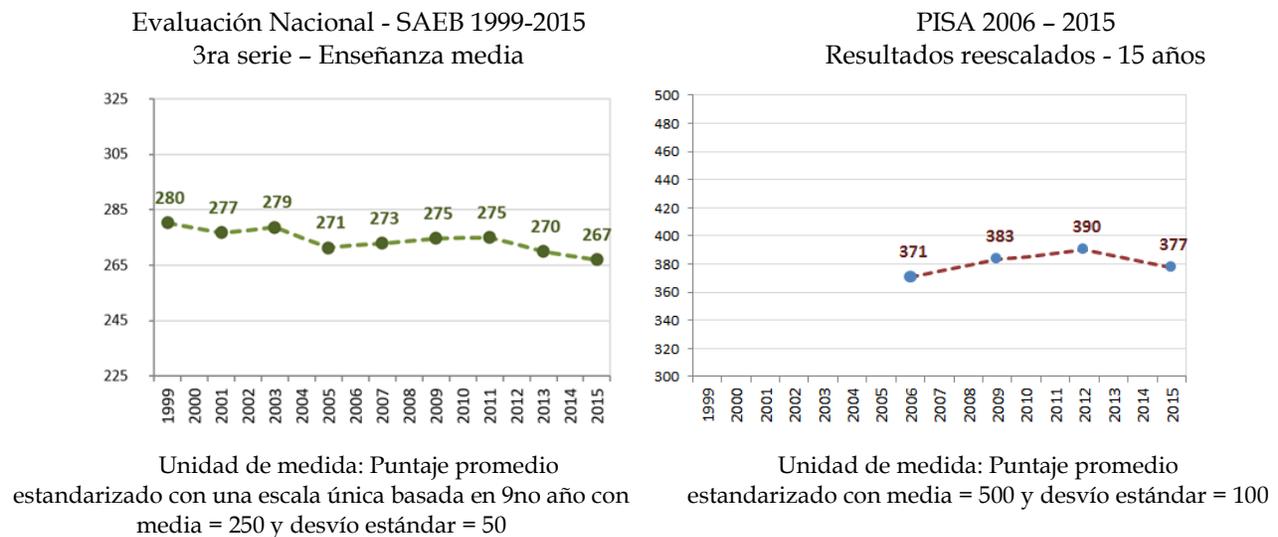
En este marco, resulta interesante resaltar que los resultados de la prueba SAEB de 3er año de la enseñanza media (edad teórica 17 años) muestran una tendencia diferente a la que perfila la finalización de la enseñanza fundamental.

Gráfico 41. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015



Fuente: CIPPEC, sobre la base de resultados SAEB 1999-2015 y OCDE-PISA 2006-2015.

Gráfico 42. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Brasil. 1999-2015

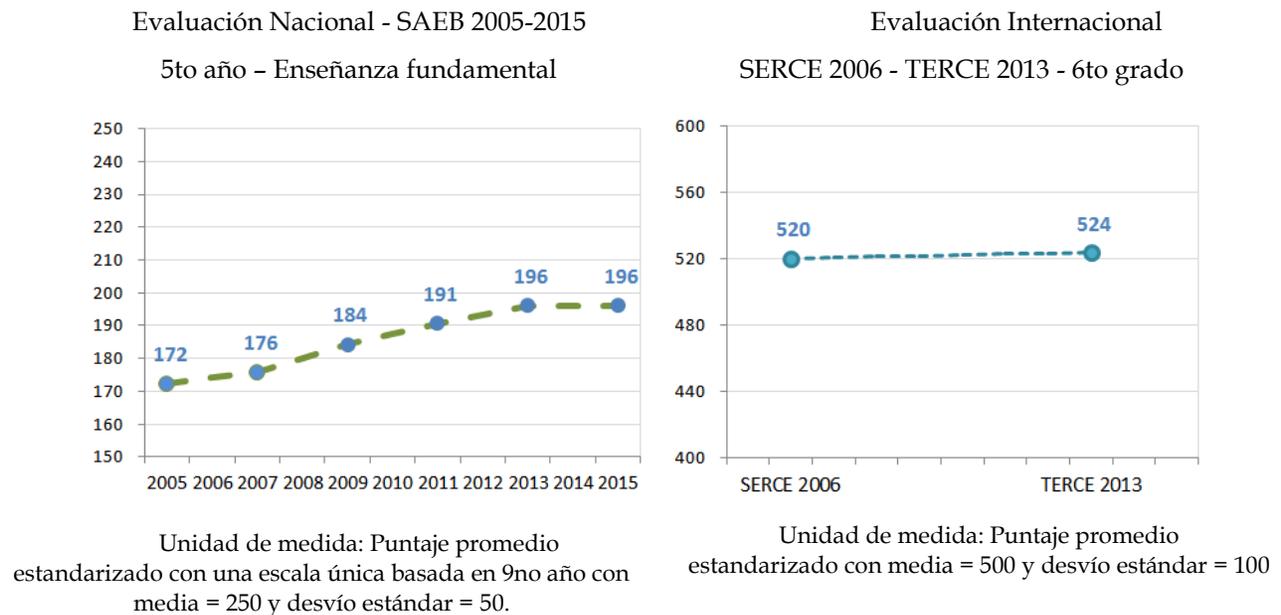


Fuente: CIPPEC, sobre la base de resultados SAEB 1999-2015 y OCDE-PISA 2006-2015.

En el área de lectura, los resultados se presentan estables en el período, con puntajes similares entre 1999 y 2015 (267 puntos). Entre los años 2003 y 2009 se concentran las principales variaciones, expresadas en una caída de los resultados en 2005 (258) y un posterior repunte. Por el contrario, los resultados de matemática para este año de estudio muestran una caída en los desempeños (280 puntos en 1999 y 267 puntos en 2015), diferencia relevante en una escala con desvío estándar de 50. En el periodo, las reducciones más marcadas se identifican en 2003 - 2005 y 2011-2015.

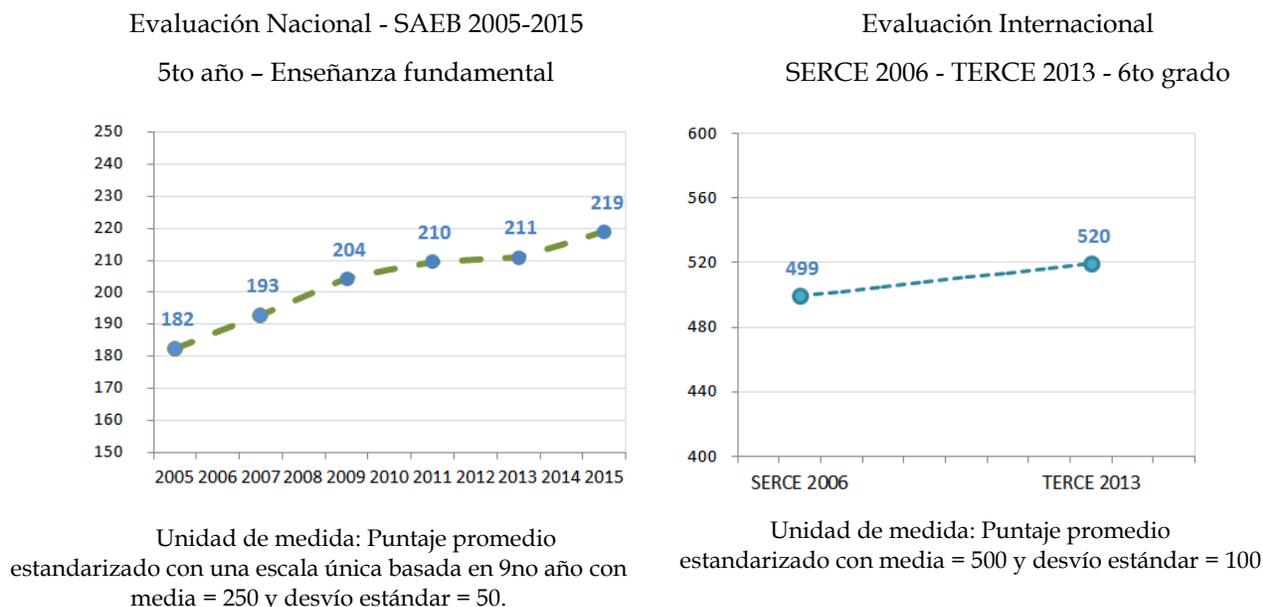
Un aspecto interesante para incorporar en el diálogo entre diferentes resultados es la tendencia de los desempeños de los estudiantes en el 5to año de la enseñanza fundamental o primaria. Las pruebas SAEB que se aplican en esta etapa muestran una mejora sostenida entre los años 2005 y 2015, con diferencias muy relevantes: un incremento de 24 puntos en lectura (lo que equivale a medio desvío estándar en la escala) y de 37 en matemática.

Gráfico 43. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Brasil. 2005-2015.



Fuente: CIPPEC, sobre la base de resultados SAEB 2005-2015 y SERCE - TERCE.

Gráfico 44. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Brasil. 2005-2015



Fuente: CIPPEC, sobre la base de resultados SAEB 2005-2015 y SERCE - TERCE.

En las pruebas TERCE del LLECE-UNESCO, los resultados de aprendizaje de nivel primario (6to grado) son útiles para contrastar esta mejora observada en SAEB. Los resultados del 2013 posicionan al país en lugares intermedios en la región, por encima del promedio de países, pero por debajo de la mayoría de los 7 países analizados en este documento. La mirada comparada SERCE-TERCE permite identificar cambios en los desempeños por área: el país incrementó sus puntajes en matemática entre 2006 y 2013 en magnitudes similares al conjunto de países: pasó de 499 a 520 puntos (+21 puntos), levemente superior a la mejora de 19 puntos en el conjunto de países. En lectura el país no evidencia cambios significativos: los puntajes apenas se incrementaron, entre ambas aplicaciones, de 520 a 524 puntos.

Es posible identificar algunos elementos de diagnóstico de las distintas evaluaciones que permiten reconocer tendencias convergentes, y algunas diferencias y particularidades. El escenario general tiende a mostrar al país con desempeños medios a bajos en relación a los 7 países seleccionados para el estudio. Por otra parte, la mayoría de las fuentes analizadas tienden a presentar un escenario de estabilidad de resultados en el nivel secundario, y ciertas mejoras en el nivel primario. Este “estancamiento” de los desempeños del nivel secundario, leído en el contexto de un incremento de la cobertura, indicaría para la OCDE el logro del país de sostener los resultados de aprendizaje en un contexto de incorporación de población históricamente excluida. Los resultados de SAEB de 3° medio parecen corroborar este escenario de estabilidad. A pesar de evaluar a una población diferente, se comportan en forma similar a PISA tanto en las tendencias históricas como en las diferencias territoriales.

A pesar de ello, la comparación de PISA con SAEB de 9° año perfila escenarios poco consistentes: mientras la evaluación nacional perfila estabilidad en la primera mitad del 2000 y luego un incremento en ambas áreas, PISA tiende a mostrar una estabilidad general de resultados en lectura, y una mejora en matemática concentrada en las primeras aplicaciones. Incluso los estudios

de Klein (2011) ponen en duda esta mejora de PISA, asociándola a cambios en la población objetivo de PISA.

Resumen de datos educativos y sociales destacados de Brasil

Cobertura

1. Los niveles históricos de cobertura de Brasil eran los más bajos entre los siete países estudiados: para 1981, la tasa neta de cobertura en el nivel primario era del 75% (SEDLAC). Incluso la tasa de analfabetismo alcanzaba, para 1980, a una cuarta parte de la población de 15 años y más (25,4%). Durante el período 1960-2014, la mejora en el acceso fue notable, y llegó a igualar los estándares de la región: La tasa neta de escolarización llega al 98,7% en 2014 en el nivel primario (SEDLAC), y al 79,5% en el nivel secundario (SITEAL). En muchos sentidos, el sistema educativo brasileiro es bastante joven en su aspiración a la universalización de la cobertura.
2. En la última década, Brasil expande el acceso a la educación secundaria, con un crecimiento que entre el año 2000 y 2014 de 9,3 puntos porcentuales, situando a la tasa neta de escolarización secundaria en un 79,5%, valor solo superior a los de Uruguay y México y por debajo del promedio de los 7 países estudiados (81,5%) con información publicada. (Circa 2014)
3. Brasil también logró grandes avances en el nivel inicial a lo largo del período estudiado: entre 2001 y 2014 la tasa bruta de matrícula de 3 a 5 años se incrementa del 45% al 73% (SEDLAC).

Trayectorias

1. La trayectoria de los alumnos atravesó una situación similar a la de la cobertura. Históricamente hubo grandes niveles de repetición: al año 2001, el 25% de los estudiantes de primaria eran repetidores, valores muy altos en relación al promedio de los 7 países (8,5%) (UIS). Sin embargo, para el año 2011 hubo una mejora los niveles de repitencia, que alcanzaron el 8,7% (UIS), aunque permanecen muy por encima del promedio de los 7 países (3,6%) (UIS- circa 2015).
2. Esa mejora no alcanzó a igualar a Brasil con el resto de los países analizados. En el año 2015, el 36,4% de los estudiantes evaluados en PISA repitieron al menos una vez de grado, este porcentaje está entre los más altos de los 7 países, superado únicamente por Colombia. Al analizar las diferencias de resultados en las evaluaciones entre los estudiantes que repitieron alguna vez y quienes no lo hicieron, se observa que, la brecha alcanza 79 puntos promedio entre lectura y matemática.

Equidad

1. En los resultados de equidad, al año 2015, Brasil tiene brechas similares al promedio de la región en resultados de aprendizaje en las pruebas PISA entre los cuartiles de mayor y menor nivel socioeconómico (84 puntos en 2015, lectura y matemática en conjunto, igual al promedio del conjunto de siete países). Adicionalmente, Brasil logró disminuir levemente la brecha entre los alumnos de mayor y menor nivel socioeconómico entre 2000 y 2015, siendo la brecha a inicios del periodo de 88 puntos.
2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que Brasil es el

tercer país con menores niveles de segregación educativa en función del nivel socioeconómico, dentro del universo de los siete países estudiados. A su vez, es el segundo país que más ha reducido los niveles de segregación tomando como punto de comparación el año 2000 (aunque partiendo de situaciones muy desiguales y persisten en ese nivel).

Contexto

1. Durante el siglo XXI se perfila un intenso mejoramiento de las condiciones de vida de la población en Brasil. Según las estadísticas oficiales del país, el porcentaje de población bajo la línea de pobreza descendió del 35,7% en 2000 al 13,3% en 2015. Un indicador comparable en la región, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, también muestra una mejora significativa en el periodo, del 69% al 51%. A su vez, el indicador de pobreza construido por PISA (índice del status económico, social y cultural, ESCS por sus siglas en inglés) perfila también una marcada mejor en las condiciones de vida de la población de 15 años entre 2000 y 2015.
2. Esta evolución de la pobreza es consistente con el comportamiento de los indicadores macroeconómicos durante la década del 2000: desde el año 2002 hasta el año 2010, el PBI se incrementó constantemente en el país, en valores constantes a 1994 (IMF-WEO), con excepción al año 2009. A partir del año 2011 comienza una desaceleración intensa en el crecimiento que se manifiesta en la situación económica más preocupante de la actualidad.
3. En estos años de crecimiento económico ha mejorado en forma marcada la distribución social de la riqueza, medida a través del coeficiente de Gini, con un importante descenso del 0,59 al 0,51 (es el segundo país con mejoras en la distribución del ingreso entre 2000 y 2012, después de Argentina). Aun así, al año 2014, Brasil tenía niveles altos de desigualdad en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados es de 0,47 (SEDLAC).
4. Al año 2014, Brasil también mostraba niveles intermedios de empleo formal (el 22% de los trabajadores no poseen cobertura de pensión, contra el 33% promedio de los siete países), niveles bajos de desempleo en 2015 (6,8% en el país y 6,7% en el promedio), y niveles intermedios de cobertura de salud (el 67,4% de la población ocupada urbana posee cobertura de salud, superior al promedio de los 7 países, de 70,4%)

Inversión educativa

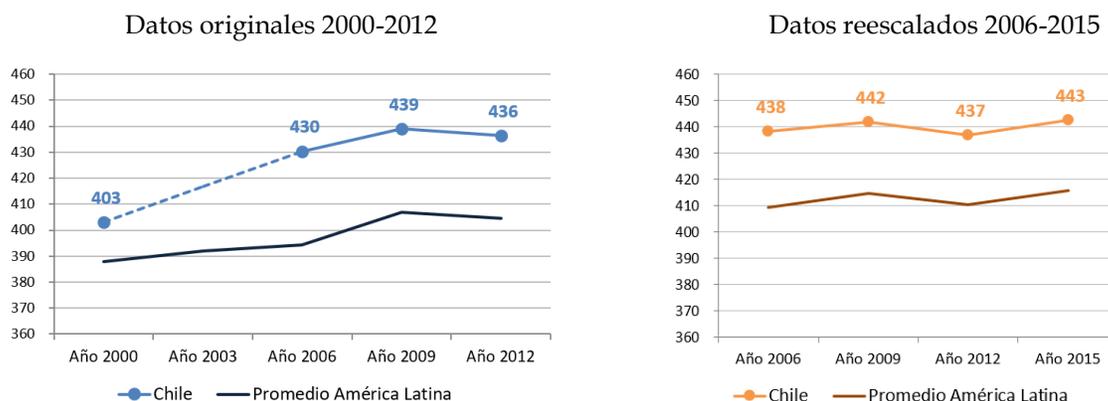
1. Brasil aumentó su inversión en el sistema educativo en el siglo XXI. Medido como porcentaje del PBI, el gasto en educación creció del 3,9% en el 2000 al 6,0% en 2013. Este crecimiento es, en puntos porcentuales, el más alto de América Latina en el periodo estudiado, después de Bolivia y Costa Rica. Gracias a este crecimiento, al año 2013 alcanza niveles de inversión (medidos como % del PBI) muy altos en la región (el promedio de 16 países de la región es 5,0% *circa* 2013).
2. Asimismo, el país destina, dentro del gasto público total, el 14,8% a educación (año 2013). En el promedio de 16 países de la región este porcentaje era del 14,3% en valores *circa* 2012. Esta porción de los recursos públicos destinados a educación mejora levemente en relación al 2000, año en el que se destinó a educación el 11,3% del gasto público.

4.3. Chile

Resultados comparados de evaluaciones de la calidad educativa

En las pruebas PISA, Chile participó en el año 2000, y en forma continua desde el 2006 al 2015. Los resultados de las pruebas PISA en Chile perfilan una situación comparada inicial por encima del promedio regional y una mejora sostenida hasta el año 2009. Al año 2000 se posicionaba entre los países de América Latina con mejores resultados en las evaluaciones, con resultados similares a México y Argentina. Los puntajes publicados originalmente en cada edición de PISA entre 2000 y 2012 muestran un incremento marcadamente superior al promedio de países (33 puntos promedio de las tres áreas en Chile, 16 puntos en el promedio de los 7 países).

Gráfico 45. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Al año 2015, es el país de la región con más altos resultados en la evaluación. En las áreas de lectura y ciencias, Chile ocupa la primera posición en América Latina desde la evaluación del año 2006.

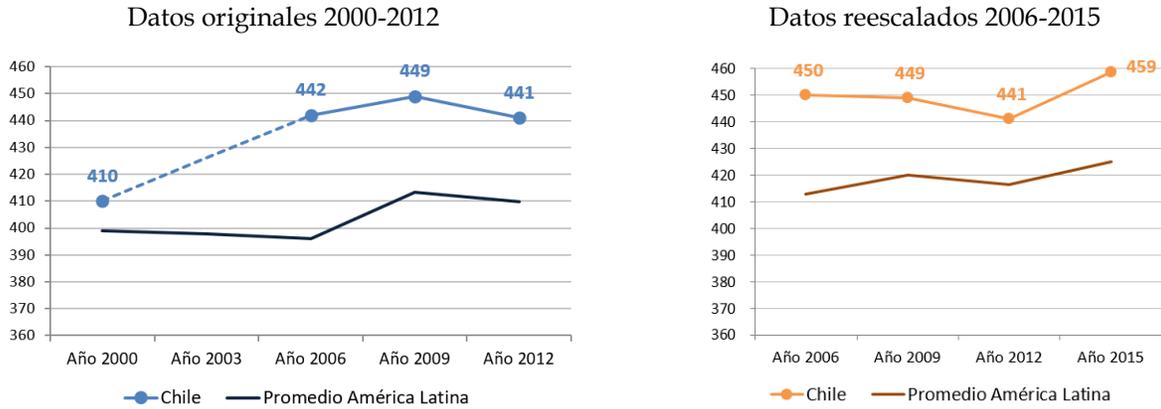
Esta significativa mejora fue menos intensa de lo que muestran estos valores si se consideran los resultados reescalados con el cambio metodológico implementado en el último ciclo de PISA, lamentablemente sólo disponibles desde el 2006 en adelante³⁶.

Los resultados publicados originales muestran una mejora promedio entre 2006 y 2015 de 12 puntos, con mayor intensidad en lectura (16 puntos) y de magnitud más baja en ciencia (9 puntos). Dados los márgenes de error de las comparaciones en el tiempo, este incremento no es estadísticamente significativo en ninguna de las tres áreas. En contraposición, según los resultados reescalados (que resultan estrictamente comparables) esta mejora promedio se reduce a 4 puntos, perfilando escasos avances en la última década. Con la metodología corregida y comparable, entre 2006 y 2015 Chile mejoró 9 puntos en lectura, 6 en matemática, y retrocedió un punto en ciencias. Ninguna de estas tres diferencias es estadísticamente significativa, lo cual muestra una estabilidad

³⁶ Para mayor detalle, ver el capítulo 2 del documento.

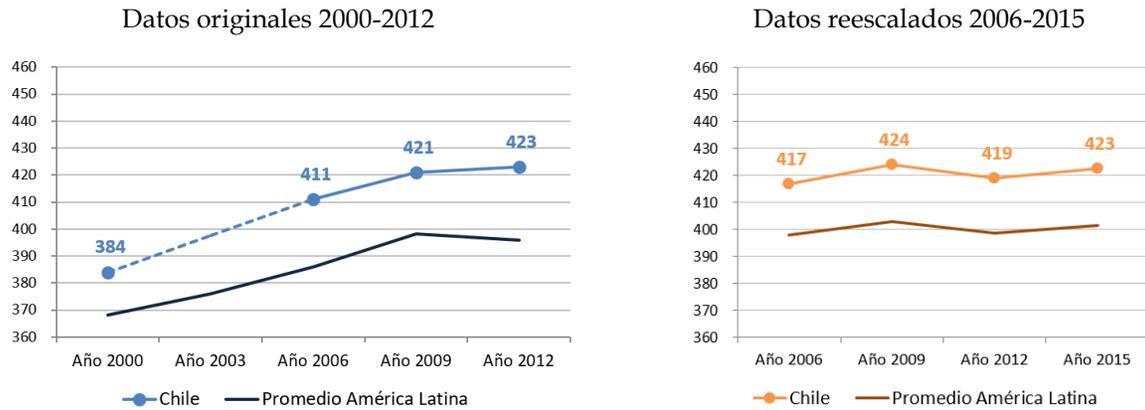
en el tiempo. No es posible estimar cuánto impactaría este cambio de metodología en la comparación de resultados PISA del año 2000.

Gráfico 46. Puntaje promedio en lectura, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Gráfico 47. Puntaje promedio en matemática, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Gráfico 48. Puntaje promedio en ciencias, resultados originales y reescalados. Chile y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

En el informe nacional de PISA 2015, Chile destaca los altos resultados obtenidos por el país y subraya la mejora observada en la evaluación de lectura respecto de 2012, en comparación con las otras dos áreas. Entre estos años se identifica un incremento de 17 puntos (Agencia de Calidad de la Educación, 2015). Esta es una “recuperación” que revierte la tendencia decreciente en el área que se observan desde 2006 según los resultados reescalados. Según los resultados comparables, en el 2015 Chile alcanza y supera los resultados previos a la caída del 2012, (aunque con diferencias no significativas).

Las evaluaciones nacionales de Chile permiten establecer una instancia de comparación con estos diagnósticos. El **recuadro 5** resume las características de su sistema nacional de evaluación de la calidad educativa.

Recuadro 5. Las evaluaciones nacionales de la calidad educativa en Chile

El país implementa las pruebas SIMCE a estudiantes del 4° y 8° año de la educación básica, y del 2° año de la educación media, con aplicaciones censales. Las evaluaciones son diseñadas para evaluar competencias, están alineadas al currículum y son referidas a criterios. La escala de puntajes se define con una media de 250 puntos y desvío estándar de 50 puntos. Chile es el país de la región que más tempranamente desarrolló evaluaciones estandarizadas de desempeño, que se implementan desde 1988, a partir de 1998 la prueba comienza a basarse en la Teoría de Respuesta al Ítem. A partir de entonces, todos los datos son comparables en el tiempo.

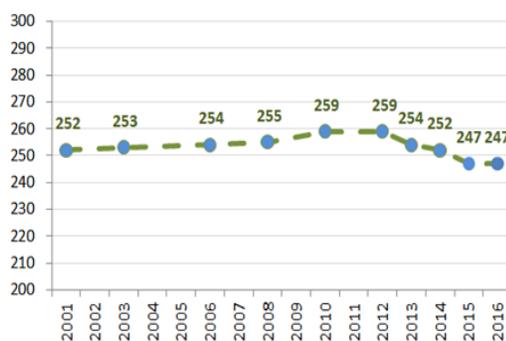
Las pruebas SIMCE evalúan el logro de los Objetivos Fundamentales y Contenidos Mínimos Obligatorios del Marco Curricular vigente en las distintas áreas curriculares evaluadas. Mientras que la medición de los puntajes de desempeño se mantiene comparable en las diferentes aplicaciones, las categorías de niveles de desempeño han sido modificadas. Hasta 2012, se utilizaron niveles de logro (con las categorías inicial, intermedio y avanzado) para describir los aprendizajes correspondientes a los distintos intervalos de puntaje, en cambio, partir de 2013, fueron reemplazados por los estándares de aprendizaje (cuyas categorías son insuficiente, elemental y adecuado). Se definen diferentes puntos de corte que determinan los niveles de desempeño (insatisfactorio, mínimo, satisfactorio y avanzado) en cada área/grado. Se evalúan las áreas de lengua y matemática en todas las aplicaciones, y se incluyeron progresivamente otras áreas y años de estudio. Asimismo, desde el 2014 el país avanzó en la medición de competencias socioemocionales, incorporando indicadores de desarrollo personal y social.

Los resultados de las pruebas SIMCE de 2° año del nivel medio permiten comparar las tendencias en términos de desempeños en lectura y matemática. La edad teórica para este grado son los 15 años de edad, lo que establece cierta equivalencia de comparación con la población objetivo de PISA.

Gráfico 49. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Chile. 2001-2016

Evaluación Nacional - SIMCE 2001-2016

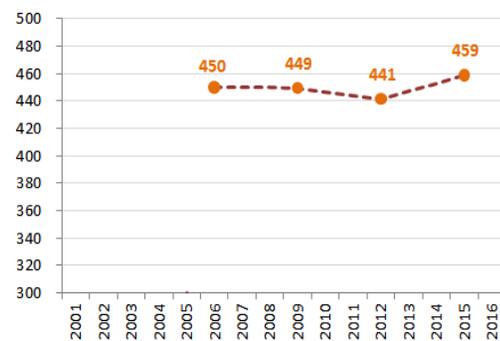
2do año educación media



Unidad de medida: Puntaje promedio

PISA 2006 - 2015

Resultados reescalados - 15 años

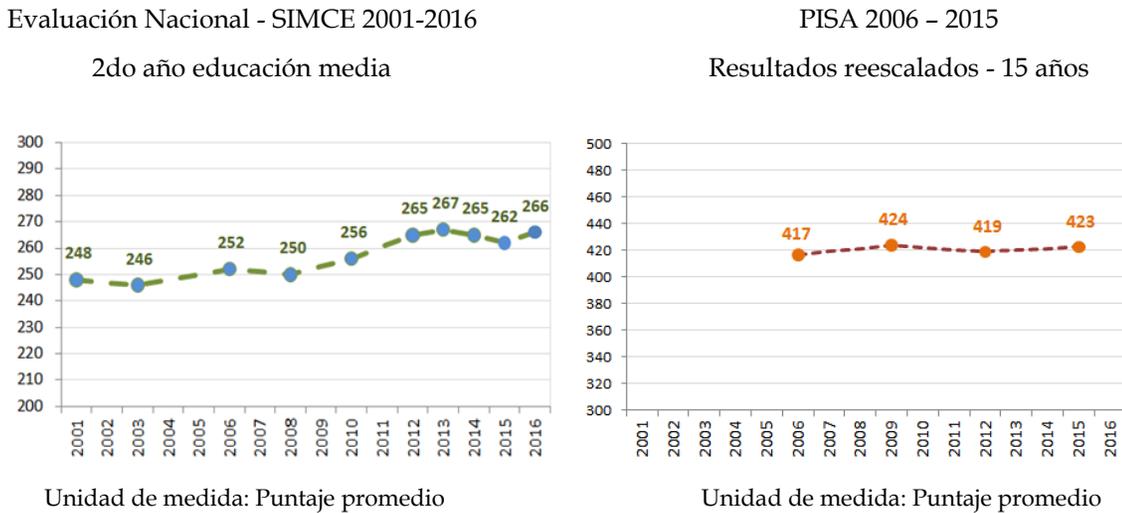


Unidad de medida: Puntaje promedio

Fuente: CIPPEC, sobre la base de resultados SIMCE 2001-2016 y OCDE-PISA 2006-2015.

En lectura, los resultados de la evaluación nacional muestran una tendencia a la mejora entre 2001 y 2010, periodo en el que los puntajes se incrementan de 252 a 259 puntos. En los años posteriores, hay un marcado cambio de sentido, los resultados de lectura tienden a una caída sostenida, alcanzando en el 2015 y 2016 los valores más bajos de las últimas dos décadas, con 247 puntos.

Gráfico 50. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Chile. 2001-2016



Fuente: CIPPEC, sobre la base de resultados SIMCE 2001-2016 y OCDE-PISA 2006-2015.

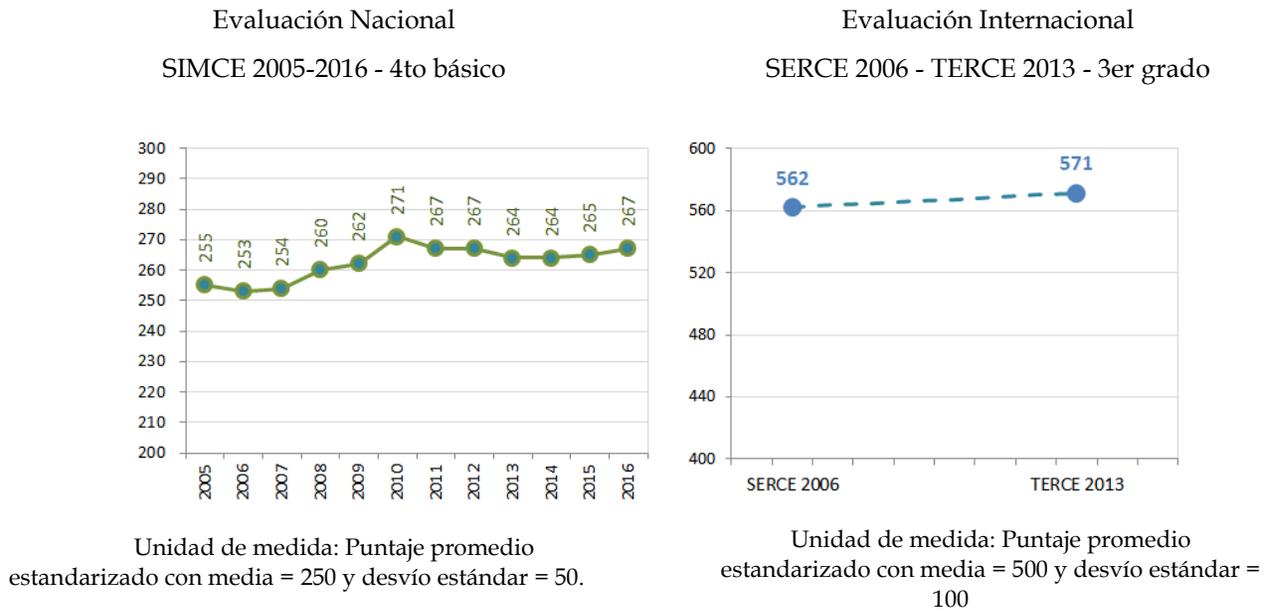
En matemática, el perfil de la tendencia es más promisorio: se observa una mejora sostenida en el período 2001 - 2016, con una diferencia neta total de 18 puntos. Se visualiza un crecimiento prácticamente constante desde el año 2001 (248 puntos) hasta el año 2013 (267 puntos). En los años posteriores el crecimiento se frena, llegando a los 266 puntos en 2016.

El comportamiento de los desempeños en las evaluaciones nacionales no muestra una correspondencia clara respecto de los resultados de PISA para el mismo período analizado. Un estudio en profundidad realizado por la Agencia de Calidad de la Educación de Chile, que analiza la concordancia entre las variaciones de las pruebas PISA y SIMCE, pone de manifiesto las divergencias en las tendencias de ambas evaluaciones en el período 2009-2012, particularmente en el área de matemática (Agencia de Calidad de la Educación, 2014).

Las evaluaciones nacionales de 4to básico (2005 - 2016) muestran un escenario relativamente similar al observado en el nivel medio. Los desempeños en lectura y matemática tienden a mostrar una mejora sostenida hasta los primeros años de la década del 2010, perfilando un escenario de avances en el período, que es seguido de un período de estancamiento relativo de los resultados. En lectura, hay una mejora marcada entre 2006 y 2010 (incremento de 253 a 271 puntos) alcanzando en este último año los valores más altos. En los años siguientes, los resultados oscilan entre los 264 y 267 puntos hasta el 2016. En matemática se observan algunas diferencias: la mejora en los desempeños se concentra entre 2007 y 2012 (246 a 261 puntos), con una posterior caída y nuevo repunte entre 2014 y 2016 (de 256 a 262 puntos). A pesar de estas diferencias, en ambos casos las variaciones de los

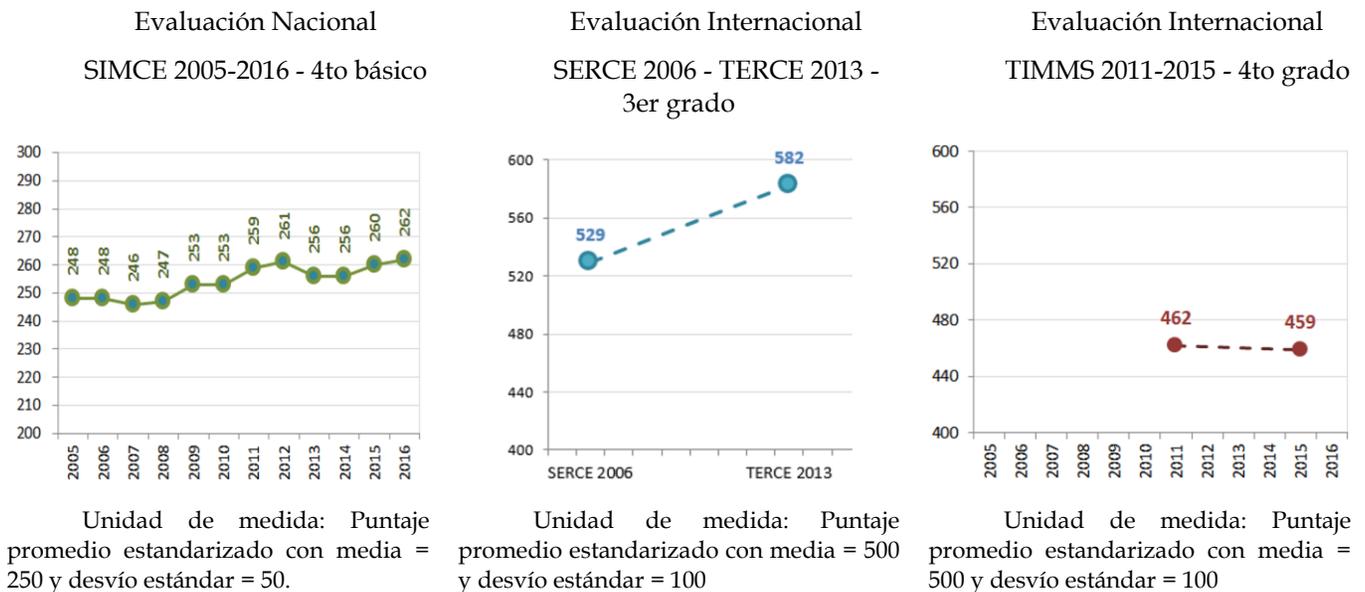
últimos años son escasas, y contrastan con el sostenido incremento de los desempeños en los años anteriores.

Gráfico 51. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Chile. 2005-2016



Fuente: CIPPEC, sobre la base de resultados SIMCE 2005-2016 y SERCE - TERCE.

Gráfico 52. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Chile. 2005-2016



Fuente: CIPPEC, sobre la base de resultados SIMCE 2005-2016 y SERCE - TERCE.

En las pruebas SERCE-TERCE del LLECE-UNESCO, los resultados de aprendizaje de nivel primario también posicionaron al país en los lugares más altos en la región, en forma similar a PISA. Ya para el SERCE (2006) de 3er grado, Chile había obtenido resultados muy por encima de la media de países evaluados, considerando en promedio los resultados de lengua y matemática (531 vs 493 en 6to). En la comparación entre SERCE y TERCE, la mejora del país en matemática fue también superior a la media, con un incremento de 53 puntos, por encima de los 31 puntos del promedio de los países. En cambio, en lectura la mejora fue inferior a la media de países, con un incremento de 9 puntos para Chile y de 18 para el promedio de países. En las evaluaciones SERCE-TERCE de 6to grado, el comportamiento entre las áreas es similar a 3er grado, y se destaca la mejora en matemática (63 puntos) muy por encima de la media de países (19 puntos) En las evaluaciones TERCE de la UNESCO, Chile fue el país con más altos resultados en todas las áreas y grados evaluados.

El Estudio de las Tendencias en Matemáticas y Ciencias (TIMMS, por sus siglas en inglés) desarrollado por la Asociación Internacional para la Evaluación del Rendimiento Educativo, construye una segunda fuente de datos internacionales para contrastar las tendencias de desempeño. Esta evaluación se orienta a los estudiantes de 4° y 8° grado, y Chile participó en las aplicaciones de 2011 y 2015 en ambos años, y también en 1999 y 2003 para 8° año. Los resultados en este último grado de la educación básica muestran una mejora sostenida, con diferencias estadísticamente significativas, en las aplicaciones 2003, 2011 y 2015 (387, 416 y 427 puntos respectivamente). En cambio, en las evaluaciones de 4° año no se han observado variaciones significativas (462 puntos en 2011 y 459 en 2015).

Es posible identificar algunos elementos de diagnóstico de las distintas evaluaciones que permiten reconocer tendencias confluyentes, y algunas diferencias y particularidades. El escenario general tiende a mostrar al país con desempeños altos y crecimiento sostenido en las últimas dos décadas, especialmente en matemática. La combinación de ambas características es un hecho destacable, debido a que las mejoras más intensas suelen concentrarse en los países más rezagados. En general las evaluaciones coinciden en señalar que el crecimiento ha sido más intenso en la década del 2000, pero hay mayores divergencias en los diagnósticos referidos a los últimos años. En algunas evaluaciones nacionales e internacionales se observa un estancamiento en la última década.

Resumen de datos educativos y sociales destacados de Chile

Cobertura

1. Chile muestra un logro histórico de cobertura: para el año 2000, la tasa neta de escolarización secundaria se situaba en el 74,1%, solamente por debajo de Argentina (SEDLAC), pero por encima del promedio de los 7 países seleccionados de la región con 71% (SEDLAC).
2. La cobertura se mantuvo en alza durante el período analizado, con mejoras moderadas considerando que el punto de partida era alto. Al 2014 la tasa de escolarización del nivel secundario alcanzó al 84,3% de la población. Este crecimiento se concentró principalmente en los sectores más desfavorecidos: para el quintil I el incremento de la tasa fue del 62% al 79% (SEDLAC).
3. También logró avances significativos en el nivel inicial a lo largo del período estudiado, especialmente en la incorporación de niños de 3 y 4 años: entre 2000 y 2013 la tasa bruta de escolarización de la población de 3 a 5 años creció del 46% al 77% para el total de población,

y alcanzando el 75% para el quintil más pobre. Esto implica también una reducción de las brechas de desigualdad en el acceso.

Trayectorias de los alumnos

1. El caso de Chile se destaca especialmente por los logros en las trayectorias educativas de los alumnos. Los niveles de repitencia en nivel primario se encuentran entre los más bajos de los 7 países. Sin embargo, en los últimos años, se han incrementado levemente. En el nivel primario, la repitencia alcanza en el año 2015 el 3,6% (UIS).
2. De acuerdo a los datos de PISA, en el año 2015, el 24,6% de los estudiantes de 15 años evaluados repitieron al menos una vez. Este valor es bajo para los niveles alcanzados en los países analizados (promedio 30%). Las brechas en los resultados de las evaluaciones de PISA Chile entre quienes repitieron y quienes no alcanzan 70 puntos en promedio entre lectura y matemática.

Equidad

1. En este contexto de mejora generalizada, Chile también manifiesta una reducción de las desigualdades en los resultados PISA: las brechas entre los cuartiles de mayor y menor nivel socioeconómico se situaron en torno a los 111 puntos en 2000, y se redujeron a 92 en el 2015. Sin embargo, en este último año Chile muestra niveles medio-altos de segregación; es el cuarto país con mayores brechas de resultados por cuartiles, detrás de Perú y Uruguay. En promedio, en el 2012 la brecha por cuartiles en los países evaluados era de 88 puntos.
2. Considerando la distribución del indicador de pobreza construido por PISA (índice del status económico, social y cultural, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que Chile es el segundo país con mayores niveles de segregación educativa en función del nivel socioeconómico dentro del universo de países estudiados. Y redujo apenas levemente sus niveles de segregación entre 2000 y 2015.

Contexto

1. Durante la última década, se perfila un mejoramiento de las condiciones de vida de la población en Chile. El porcentaje de población bajo la línea de pobreza descendió del 20% en 2000 al 12% en 2015. Un indicador comparable en la región, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, también muestra una mejora significativa en el periodo, del 35% al 20% (2013). A su vez, el indicador de pobreza construido por PISA (índice del status económico, social y cultural, ESCS por sus siglas en inglés) perfila también una marcada mejora en las condiciones de vida de la población de 15 años entre 2000 y 2015. Según esta fuente, la población adolescente en Chile se encuentra en mejores condiciones que en el resto de los países.
2. Esta evolución de la pobreza es consistente con el comportamiento de los indicadores macroeconómicos: el inicio de la década del 2000 está marcado por un crecimiento sostenido. Entre los años 2000 y 2014 el PBI creció a un promedio del 4,1% anual. En este periodo, el año 2009 muestra un contexto de crisis, con una contracción del PBI per cápita del 1,4%. Para el año final de la serie, es el país con más alto PBI per cápita de los siete analizados (IMF-WEO).
3. Entre los años 2000 y 2014, se observa también una mejora en la distribución social de la riqueza, medida a través del coeficiente de Gini, con un leve descenso del 0,55 al 0,50 al 2013.

Sin embargo, Chile pertenece, junto con Colombia y Brasil, al grupo de países con mayores niveles de desigualdad en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados es de 0,47 (SEDLAC).

4. Chile también manifiesta muy altos niveles de empleo formal (el 15,5% de los trabajadores no poseen cobertura de pensión, contra el 33,7% promedio de los siete países), niveles bajos de desempleo (6,2% en el país y 6,6% en el promedio), y muy altos niveles de cobertura de salud (el 95% de la población ocupada urbana posee cobertura de salud, la segunda más alta de los 7 países estudiados, cuyo promedio es de 70%)

Inversión educativa

1. La inversión educativa en el país, medida como porcentaje del PBI, aumentó marcadamente en la segunda mitad de la década del 2000, pero luego de un periodo de decrecimiento. Creció del 3,02% en el 2006 al 4,24% en 2009, y luego se mantuvo estable, alcanzando en 2014 el 4,7%. Sin embargo, si se considera el periodo completo 2000 - 2012, la intensidad del crecimiento se reduce, ya que sólo se incrementó del 3,7% al 4,7% del PBI. Por esta razón, al año 2012 pertenece al grupo de países con niveles medio-bajos de inversión (medidos como % del PBI) (el promedio de 16 países de la región es 5,0% en valores *circa* 2013).
2. Asimismo, el esfuerzo de inversión pública total (medido como gasto público como porcentaje del PBI) muestra un muy bajo incremento entre 2000 y 2014, en torno a 1 punto porcentual (el tercer país con menor crecimiento en el periodo). Cabe señalar que el gasto público disminuye fuertemente frente al PBI hasta el año 2006, donde se sitúa en el 18,7%, seguido de años de crecimiento. Al año 2014, el gasto público equivalía al 24,0% del PBI, lo que posicionaba al país en el 10mo lugar en el conjunto de países de América Latina, cerca de Paraguay (23,5%) y Panamá (23,6%).

4.4. Colombia

Resultados comparados de evaluaciones de la calidad educativa

En las pruebas PISA, Colombia participó en los años 2006 a 2015, con lo cual su evolución es más difícil de juzgar porque es más corta. Otro elemento desafiante para el análisis es la diferencia de diagnóstico al comparar los resultados publicados originalmente por PISA y los reescalados con el cambio metodológico implementado en el último ciclo de PISA³⁷.

Gráfico 53. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015

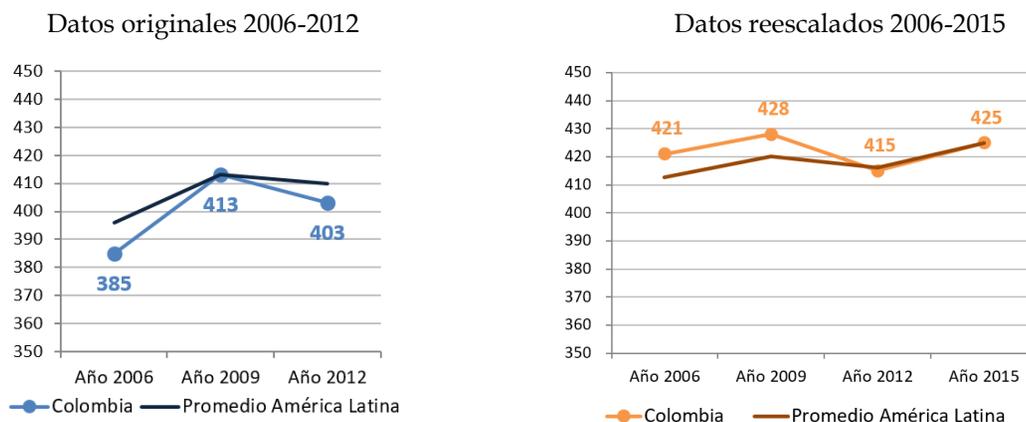


Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2006-2015.

Con la metodología utilizada en las publicaciones oficiales (que no es estrictamente comparable), Colombia tendría una marcada mejora en sus resultados, con un incremento estadísticamente significativo de 29 puntos (promedio de las tres áreas). Pertenecería al grupo de países con mayor crecimiento en el período. Sin embargo, esta mejora se ve puesta en cuestión cuando se analiza la evolución con los resultados reescalados: entre 2006 y 2015, los desempeños de Colombia se mantienen estables (una mejora promedio de 3 puntos, no significativa en términos estadísticos), dando cuenta de una situación de escasos avances en la última década. Algunos ejemplos permiten ilustrar estas diferencias.

³⁷ Para mayor detalle, ver el capítulo 2 del documento.

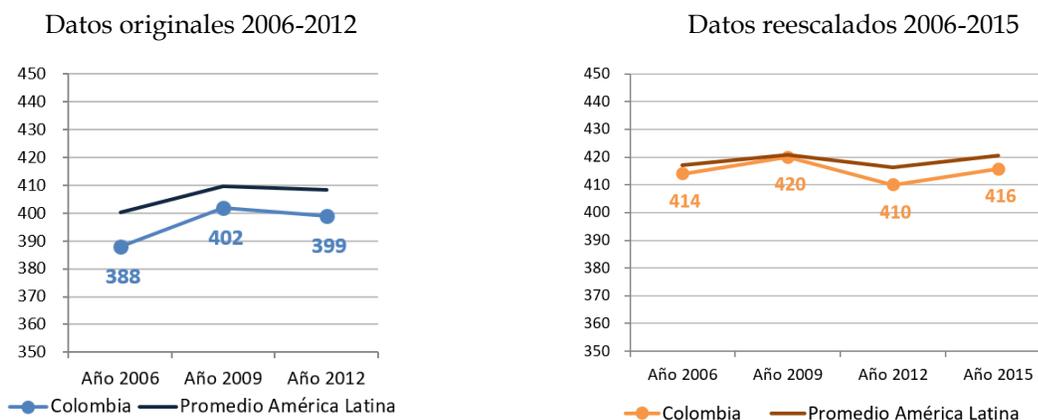
Gráfico 54. Puntaje promedio en lectura, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2006-2015.

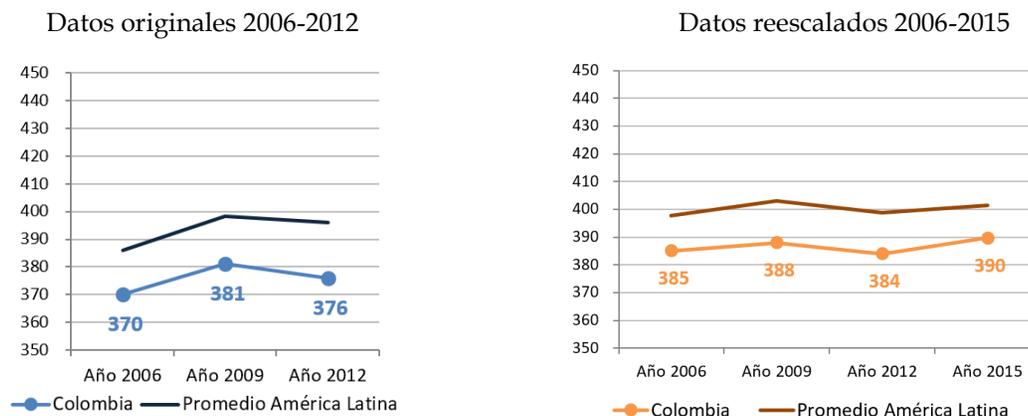
Según el informe nacional de PISA 2015, Colombia manifiesta en el área de lectura el mayor progreso, con una diferencia de 40 puntos respecto de 2006. Con un incremento de 385 a 425 puntos, es el cuarto país que más progresó en el área en el período. Los resultados obtenidos son interpretados como un significativo avance en la reducción de las brechas de desigualdad en el aprendizaje (ICFES, 2016). Ahora bien, si se utiliza la metodología de escalamiento de PISA 2015 que permite realizar una comparación estricta con años anteriores, Colombia hubiese obtenido 421 puntos en lectura en 2006. Es decir, la mejora se reduciría a unos escasos 4 puntos, dato que pone en duda el avance del país en sus desempeños.

Gráfico 55. Puntaje promedio en ciencias, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2006-2015.

Gráfico 56. Puntaje promedio en matemática, resultados originales y reescalados. Colombia y promedio de países seleccionados de América Latina. 2006-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2006-2015.

Por su parte, la OCDE, en sus notas destacadas del país, resalta los resultados de ciencias obtenidos en Colombia, indicando que incrementó sus desempeños en 28 puntos entre 2006 y 2015, la segunda mayor mejora entre los 52 países con resultados comparables en el periodo (OCDE, 2016d). También esta mejora sustantiva queda puesta en cuestión con los reescalamientos de resultados que PISA realiza a partir de los cambios metodológicos implementados en 2015: la diferencia de puntajes se reduciría a 2 puntos si se calcula el resultado del 2006 con la nueva metodología.

Según se ha explicado en el capítulo 2, este incremento de los desempeños evidenciado en los resultados originales de cada aplicación de PISA se vincula con los altos porcentajes de *non-reached items* en las primeras participaciones del país. Hasta la aplicación 2012 inclusive, en promedio los estudiantes dejaban en blanco más del 10% de los ítems finales de la evaluación. La OCDE afirma en el informe de PISA 2015 que “la mejora en el rendimiento medio de ciencias reportada para Colombia se debe casi totalmente a cambios en el enfoque a la escala. El aumento de la puntuación media habría sido de sólo tres puntos (no significativos) si se hubiera utilizado el enfoque de 2015 y la muestra de calibración para escalar los resultados de 2006” (OCDE, 2016).

Las evaluaciones nacionales de Colombia permiten establecer una instancia de comparación con estos diagnósticos divergentes. El **recuadro 6** resume las características de las principales evaluaciones nacionales de la calidad educativa implementadas en Colombia.

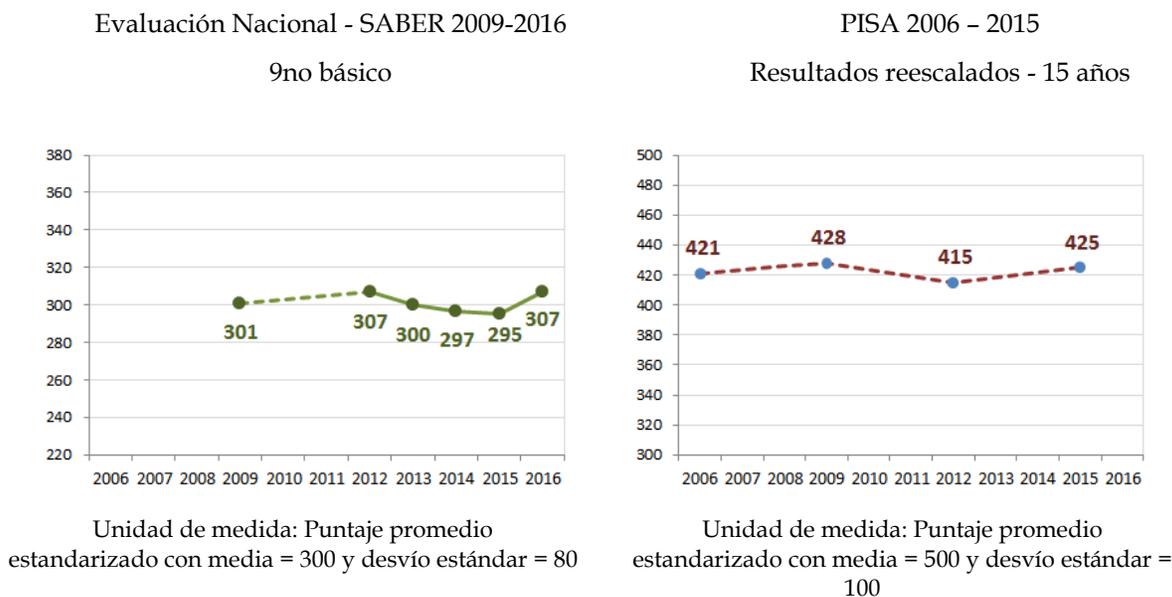
Recuadro 6. Las evaluaciones nacionales de la calidad educativa en Colombia

Colombia aplica las pruebas SABER en forma censal a estudiantes del 3°, 5°, 9° y 11° año de estudio. La institución responsable de su implementación es el Instituto Colombiano para la Evaluación de la Educación (ICFES). Se aplicaban con una frecuencia trienal hasta 2012, cuando comenzaron a aplicarse anualmente.

Las evaluaciones son diseñadas para evaluar competencias, están alineadas al currículum y son referidas a criterios. Se evalúan las áreas de lengua y matemática en todas las aplicaciones, se aplica en forma alterna una evaluación de ciencias naturales o pensamiento ciudadano. También se incluyen en SABER 11° otras áreas: biología, física, química, filosofía e inglés.

La escala de puntajes va desde los 100 a los 500 puntos, se define con una media de 300 puntos y un desvío estándar de 80 puntos, y es comparable desde el año 2009 en adelante (excepto en SABER 11°, cuya escala tiene media 50 y desvío 10, y el período de comparabilidad es mayor). Se definen diferentes puntos de corte que determinan los niveles de desempeño (insatisfactorio, mínimo, satisfactorio y avanzado) en cada área/grado.

Gráfico 57. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016



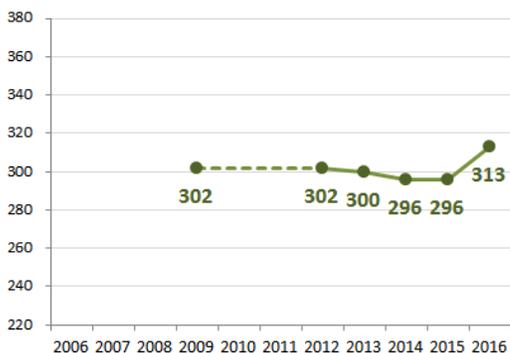
Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y OCDE-PISA 2006-2015.

Los resultados de las pruebas SABER 9 muestran resultados estables en las tres áreas (lectura, matemática y ciencias). La edad teórica para este grado son los 14 años de edad, lo que establece cierta equivalencia de comparación con la población objetivo de PISA (15 años de edad). En lectura, los resultados entre 2009 y 2016 se incrementaron de 301 a 307 puntos, lo que constituye una diferencia muy pequeña si se considera que el desvío estándar de la escala es 80 puntos.

Gráfico 58. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016

Evaluación Nacional - SABER 2009-2016

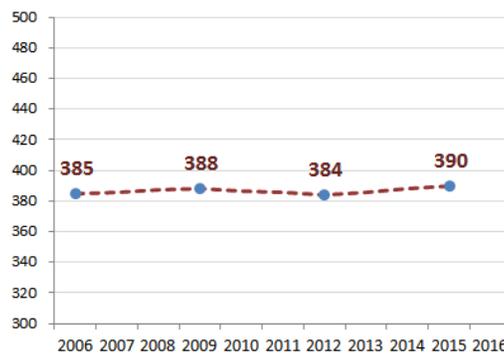
9no básico



Unidad de medida: Puntaje promedio estandarizado con media = 300 y desvío estándar = 80

PISA 2006 - 2015

Resultados reescalados - 15 años



Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y OCDE-PISA 2006-2015.

Gráfico 59. Resultados en ciencias en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016

Evaluación Nacional - SABER 2009-2016

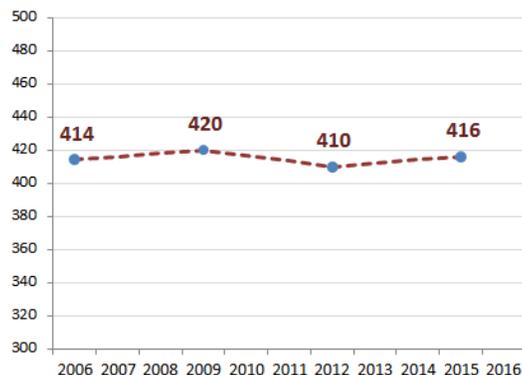
9no básico



Unidad de medida: Puntaje promedio estandarizado con media = 300 y desvío estándar = 80

PISA 2006 - 2015

Resultados reescalados - 15 años

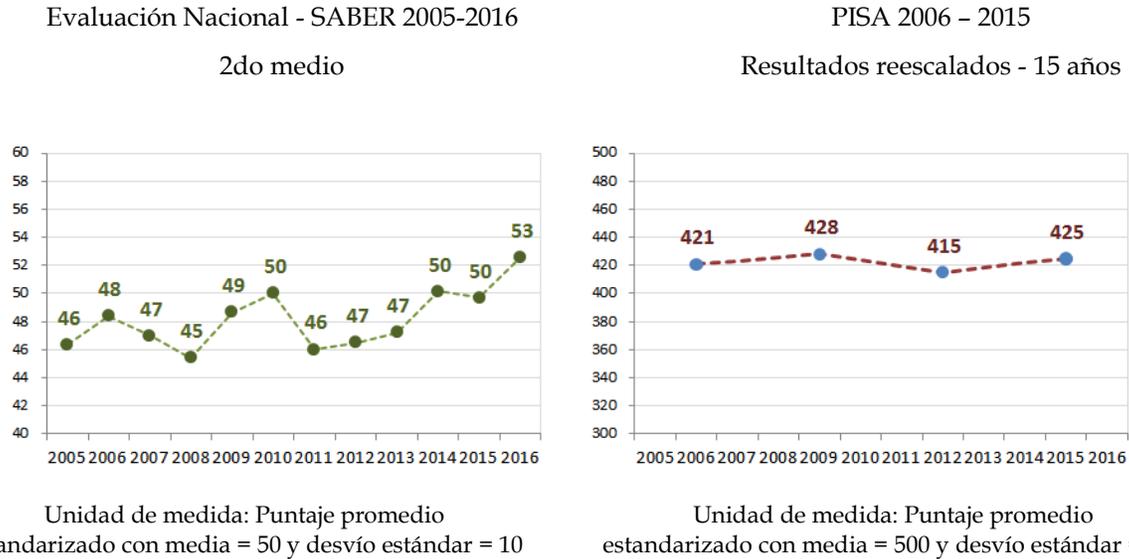


Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y OCDE-PISA 2006-2015.

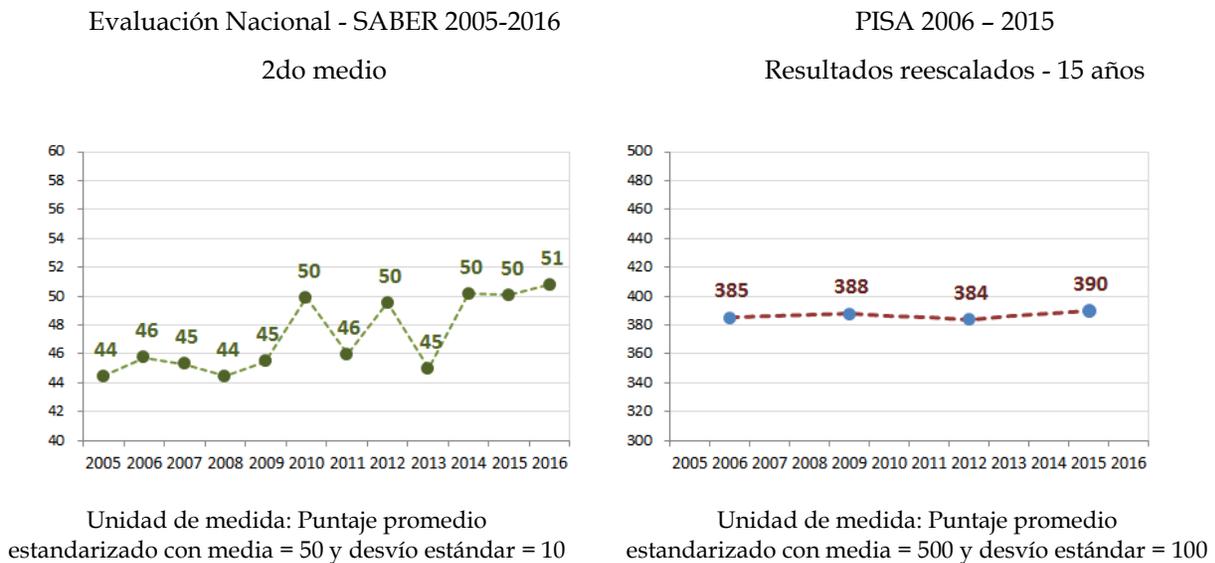
En matemática se observa una mejora también leve, de 11 puntos, en el período. Con un resultado de 302 puntos en 2009, el incremento se concentra exclusivamente entre 2015 y 2016, donde el puntaje aumenta de 296 a 313 en sólo un año. Ciencias es la única área que muestra un comportamiento sostenido hacia la caída desde el año 2012 en adelante, período en el cual los resultados van de 312 a 289 puntos.

Gráfico 60. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2005-2016



Fuente: CIPPEC, sobre la base de resultados SABER 2005-2016 y OCDE-PISA 2006-2015.

Gráfico 61. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2005-2016

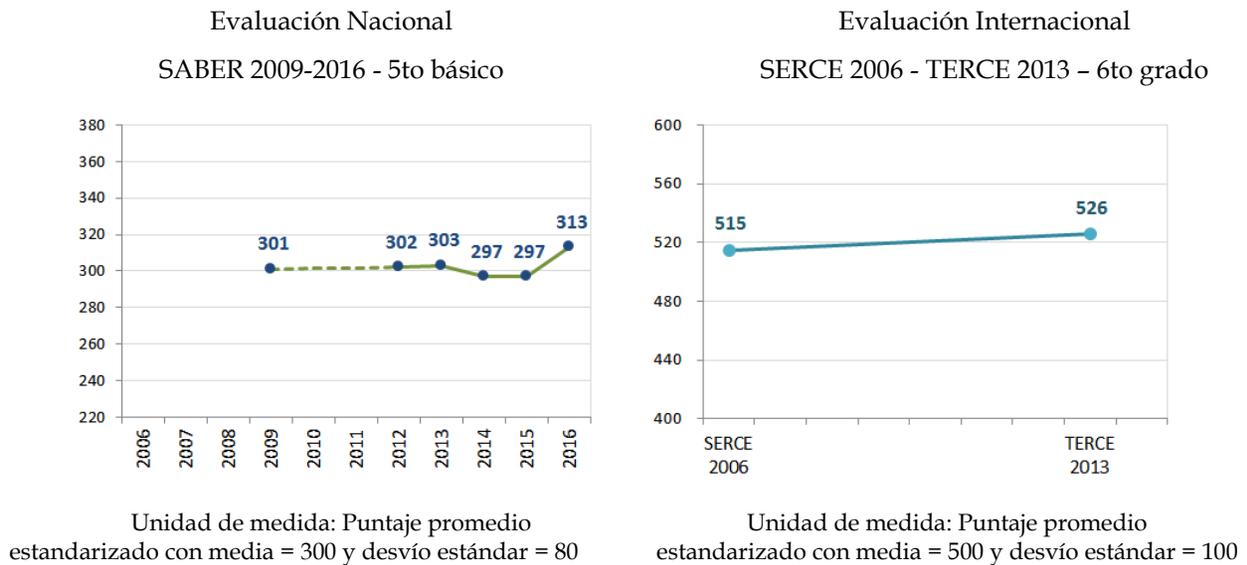


Fuente: CIPPEC, sobre la base de resultados SABER 2005-2016 y OCDE-PISA 2006-2015.

La prueba SABER 11 del 2do año del nivel medio (año de estudio cuya edad teórica son los 16 años) muestra un comportamiento diferente. En el área de lectura se identifica una sistemática tendencia a la mejora en los desempeños desde el año 2011 al 2016. En el período, los resultados se incrementan de 46 a 53 puntos, una diferencia considerable teniendo en cuenta que el desvío estándar de la escala es de 10 puntos. En matemática, los desempeños muestran una tendencia a la mejora en la mirada de largo plazo (mejora de 44 a 51 puntos entre 2005 y 2016), pero un estancamiento desde el 2010, donde los resultados se mantienen en torno a los 50 puntos. Un elemento a tener en cuenta en la comparación entre SABER 9 y SABER 11 es la diferencia en la población evaluada, considerando el abandono que ocurre entre un año de estudio y otro. El porcentaje de población que asiste a la escuela en la edad teórica constituye una aproximación a este abandono, que al año 2016 disminuye entre los 14 y 16 años del 93% al 82%³⁸.

Las evaluaciones nacionales de 5to básico –que evalúan a los estudiantes de nivel primario– muestran un escenario similar al observado en SABER 9. Los desempeños en lectura, matemática y ciencias tienden a comportarse en forma estable entre 2009 y 2014, perfilando un escenario de pocos avances en el período (cambios de 301 a 297 puntos en lectura, de 301 a 292 en matemática, y de 302 a 307 en ciencias). Ahora bien, en los tres casos se reconocen mejoras en los desempeños concentradas al final del periodo. En matemática la mejora se visualiza en los dos últimos años, los resultados alcanzan los 305 puntos para el 2016. En lectura la mejora se sitúa sólo en 2015-2016, con un incremento de 16 puntos en sólo un año. En el área de ciencias, la mejora es de 11 puntos, llegando a 318 en 2016.

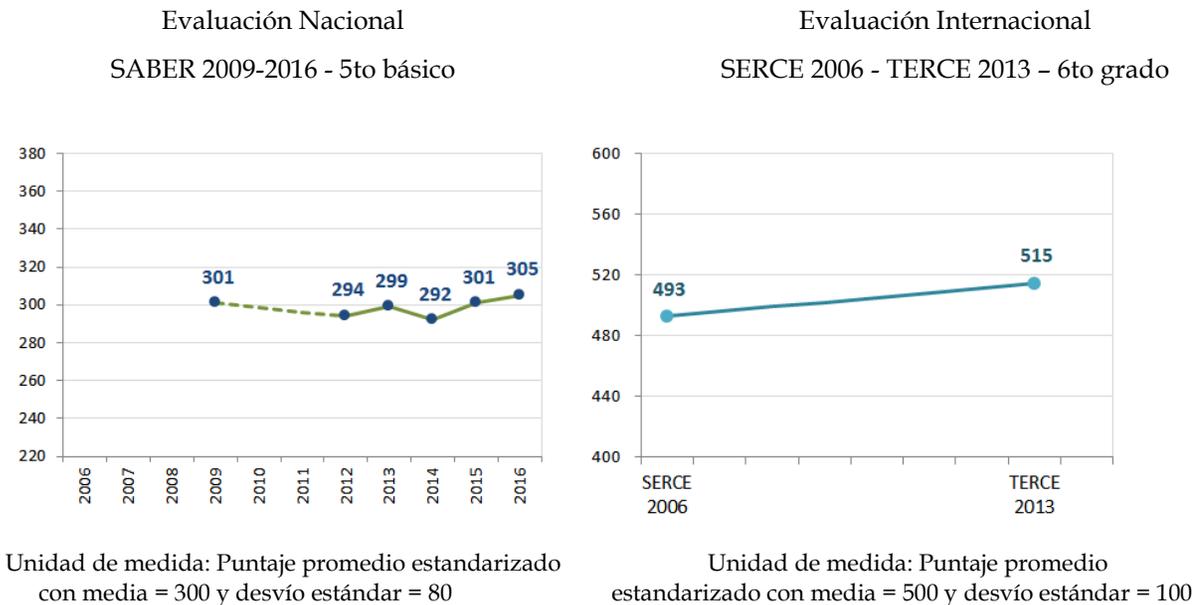
Gráfico 62. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Colombia. 2006-2016



Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y SERCE - TERCE.

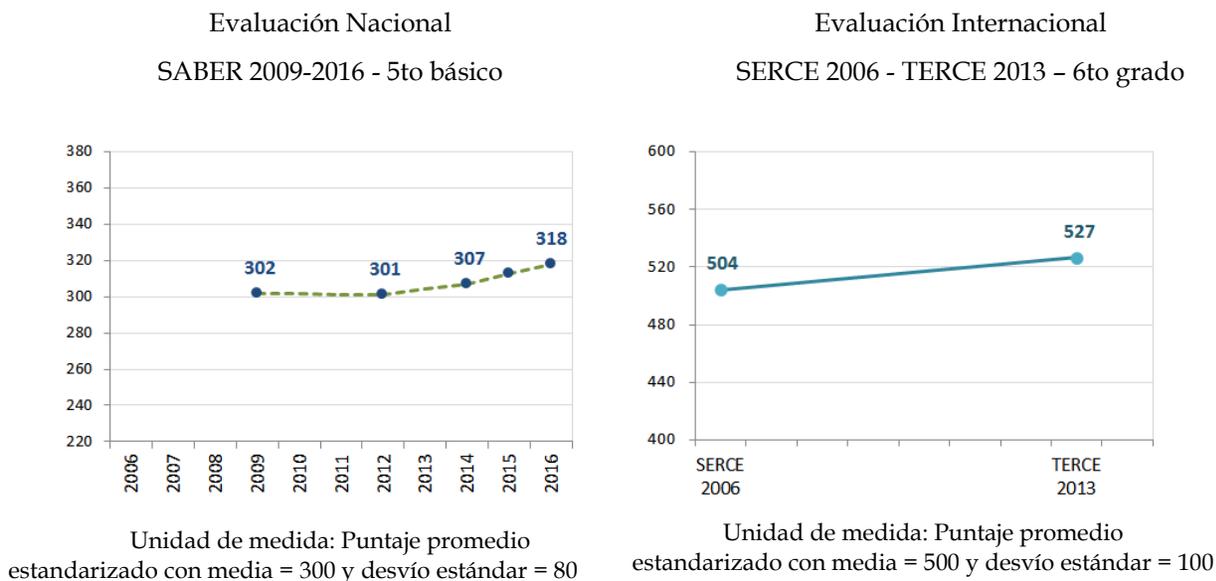
³⁸ Procesamientos propios en base a microdatos de la Gran Encuesta Integrada de Hogares (GEIH) 3er trimestre 2016, Departamento Administrativo Nacional de Estadística (DANE), Colombia.

Gráfico 63. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Colombia. 2006-2016



Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y SERCE - TERCE.

Gráfico 64. Resultados en ciencias en pruebas nacionales e internacionales. Nivel secundario. Colombia. 2006-2016



Fuente: CIPPEC, sobre la base de resultados SABER 2009-2016 y SERCE - TERCE.

En las pruebas SERCE-TERCE, los resultados de aprendizaje de nivel primario posicionaron al país en lugares intermedios. Los resultados del país son muy similares al promedio de países evaluados: 520 puntos promedio lectura y matemática en 6to grado (en la región el puntaje promedio fue de 509). En la comparación en el tiempo, el país se comportó en forma similar a la tendencia del conjunto: en promedio, los resultados de lengua y matemática mejoraron en 14 puntos en 3er grado, y 16 en 6to. En el total de países evaluados, esta tendencia fue positiva en ambas evaluaciones, con diferencias de 25 y 16 puntos, respectivamente. Colombia se destaca por haber obtenido altos resultados en la evaluación de ciencias: con un puntaje promedio de 527 se posiciona como el 4to país con mejores resultados, y de los 7 evaluados en ambos estudios ERCE, Colombia muestra un incremento de 23 puntos, el segundo más importante.

Resulta complejo obtener conclusiones confluyentes que sintetizen las tendencias perfiladas por las diferentes evaluaciones. El diagnóstico más recurrente es el de un escenario de estabilidad o de crecimiento leve en los desempeños de los estudiantes en los últimos 10 años. Ciertas tendencias contrapuestas arrojan ciertas dudas sobre cuáles son las áreas que se han visto más fortalecidas en el período. Los resultados de PISA con el reescalamiento parecen consistentes con las evaluaciones nacionales al mostrar una tendencia de estancamiento en los aprendizajes medidos por las evaluaciones estandarizadas de la calidad.

Resumen de datos educativos y sociales destacados de Colombia

Cobertura

1. Del conjunto de países estudiados, Colombia es uno de los que presenta los mayores déficits históricos de escolarización: para 1980, el 18,6% de la población joven y adulta (15 años y más) era analfabeta, el segundo valor más alto luego de Brasil (SEDLAC). La Tasa Neta de Escolarización Primaria recién logra superar la barrera del 95% hacia el año 2003 (SEDLAC)
2. En la educación secundaria, para el año 2001, la tasa neta de escolarización secundaria se situaba en el 67,7% (SEDLAC), solamente por encima México, y por debajo del promedio de los siete países analizados (71,4).
3. Desde este punto de partida en la última década muestra una amplia expansión de la educación secundaria: entre el 2000 y el 2014, la tasa neta de escolarización del nivel se incrementó 12,9 puntos. Este marcado crecimiento es distintivo del país, solo superado por Perú y México. Tanto es así que, para el año 2014, Colombia manifiesta indicadores de cobertura superiores a Brasil (79,5%) y a Uruguay (80,6%), países que tenían mayores niveles de cobertura al inicio de la década. Esta expansión benefició a todos los sectores sociales, pero se vieron particularmente favorecidos los adolescentes de los sectores más pobres: para el quintil I, el incremento de la tasa fue del 55% al 72% (SEDLAC).
4. Colombia aún manifiesta un importante retraso en el acceso a la educación inicial. A pesar de haber expandido su cobertura en los años recientes (la tasa bruta de escolarización de la población de 3 a 5 años se incrementa el 51% al 59% entre 2006 y 2014), en este último año es la más baja del conjunto de países, con una importante brecha en relación al resto (para los 7 países, el promedio del indicador es del 76%).

Trayectorias de los alumnos

1. En Colombia, las tasas de repitencia en el nivel primario se encontraban, en el 2015, en niveles bajos en relación al resto de países (2,3%), segundo país con valores más bajos después de

México y Argentina. Durante el período estudiado, se manifiesta una sensible mejora en las trayectorias del nivel primario (al año 2001 era de 5,4%).

2. La dinámica de las trayectorias escolares también puede perfilarse a partir de los cuestionarios de PISA: de los siete países estudiados es el que tiene mayor porcentaje de estudiantes que afirman haber repetido al menos una vez. El 46,2% de los estudiantes de 15 años encuestados afirma que al menos una vez repitió del grado, un porcentaje superior al promedio de países (30,5%)

Equidad

1. Observando la equidad de los desempeños según PISA, las brechas en los resultados entre los cuartiles de mayor y menor nivel socioeconómico se situaron en torno a los 76 puntos en el año 2006, y se mantuvieron a lo largo del período analizado. Al año 2015, en comparación con el resto de los países, las brechas son reducidas, ocupando un segundo lugar después de México.
2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que Colombia es, del conjunto de los siete países analizados, el tercer país con mayores niveles de segregación educativa en función del nivel socioeconómico.

Contexto

1. El indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) perfila una mejora en las condiciones de vida de la población escolar de 15 años entre 2006 y 2015 (el indicador aumenta de -1,44 a -0,98). Otros indicadores internacionales de medición de la pobreza muestran mejoras: por ejemplo, el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC se reduce de 52,7% en 2008 a 48,3% en 2014. También las estadísticas oficiales de pobreza indican una marcada mejoría, lo que corrobora las tendencias antes expuestas: según estas el porcentaje de población bajo la línea de pobreza descendió del 45,0% al 27,8% entre 2006 y 2015, y el de pobreza extrema se redujo del 10,4% al 7,9% entre 2012 y 2015.
2. Esta evolución de la pobreza puede vincularse con el comportamiento de los indicadores macroeconómicos: entre el 2006 y el 2014 el PBI per cápita se incrementó un 48% a valores de paridad de compra (IMF-WEO). Este valor sitúa a Colombia en un rango intermedio de crecimiento, en relación con el conjunto de países analizados. El incremento promedio del PBI entre 2006 y 2014 para los 7 países es del 45,9%
3. El crecimiento anual del PBI, a precios constantes, se posiciona en torno al 4,7% entre los años 2006 y 2014 (IMF-WEO), aunque se identifica una marcada variabilidad en los períodos: mientras que los años 2006 y 2007 están marcados por un crecimiento anual del PBI cercano al 7%, en los años siguientes se reconoce una intensa desaceleración del crecimiento económico que llega al 1,6% de crecimiento anual en 2009. Para los años más recientes, el PBI vuelve a crecer a ritmos menores pero más estable, en torno al 4,5% desde el 2012 en adelante.
4. En estos años se mantuvo relativamente estable la distribución social de la riqueza, medida a través del coeficiente de Gini, con un leve descenso del 0,56 al 0,53 entre 2008 y 2014. Colombia es, entre los países estudiados, el que manifiesta mayores niveles de desigualdad

en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados, para el año 2014 es de 0,47 (SEDLAC).

5. Colombia también manifiesta niveles intermedios de empleo informal (el 38% de los trabajadores no posee cobertura de pensión, contra el 33% promedio de los siete países). El porcentaje de trabajadores en empleo informal disminuyó más de veinte puntos en la última década. Los niveles de desempleo son los más altos entre los siete casos estudiados (8,9% en el país y 6,6% en el promedio), y son muy bajos sus niveles de cobertura de salud (el 48% de la población ocupada urbana posee cobertura de salud; junto con México, Colombia es de los países con más bajos niveles de cobertura de salud, cuyo promedio es de 70,4%).

Inversión educativa

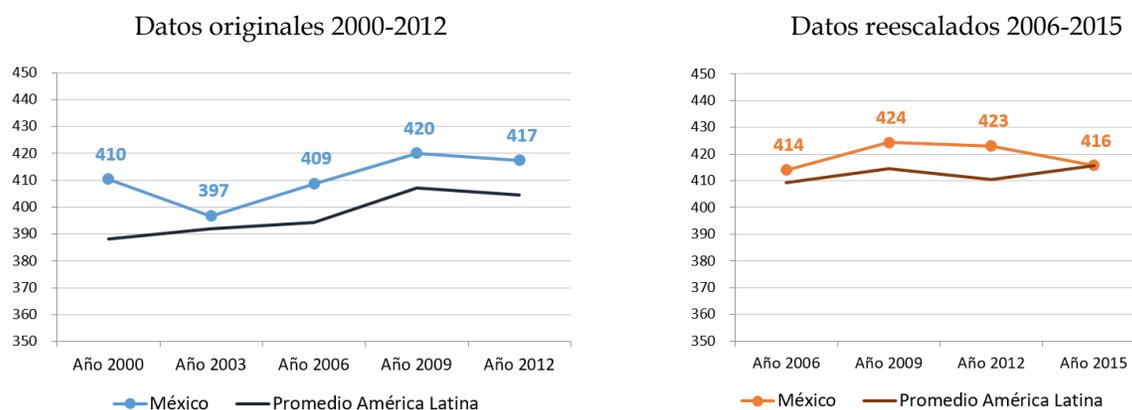
1. El país aumentó su inversión en el sistema educativo en la última década, pero comparativamente con el resto de la región lo hizo en forma moderada. Medido como porcentaje del PBI, el gasto en educación creció del 3,5% en el 2000 al 3,9% en 2006 y 4,6% en 2014. Este crecimiento es, en puntos porcentuales, levemente inferior al crecimiento promedio de los países de América Latina en el periodo (1,2 puntos, circa 2000-2013). Gracias a este crecimiento, al año 2014 alcanza niveles de inversión (medidos como porcentaje del PBI) en un rango intermedio en la región (el promedio de 16 países de la región es 5,0% en valores circa 2013).
2. Este incremento se desarrolla en un contexto donde el esfuerzo de inversión pública total permanece relativamente estable entre 2000 y 2012, lo que pone de relieve la importancia del incremento del financiamiento educativo. En Colombia, entre 2000 y 2012 el gasto público total, medido como % del PBI, creció sólo 1,8 puntos (uno de los países con menor crecimiento en el periodo). Al año 2014, el gasto público equivalía al 29,5% del PBI, lo que posicionaba al país en lugares intermedios en el conjunto de países de América Latina, donde el promedio simple del indicador es también del 29,5% para ese mismo año.

4.5. México

Resultados comparados de evaluaciones de la calidad educativa

En las pruebas PISA, México participó en forma ininterrumpida desde el año 2000 al 2015 en las sucesivas aplicaciones. Los resultados de las pruebas del país publicados por la OCDE perfilan un escenario global de estabilidad, con variaciones que van de los 410 puntos en el 2000 a los 417 en el 2012 (promedio 3 áreas). En un escenario donde varios países mejoraron sus resultados, México pasó de manifestar desempeños altos para la región al inicio de PISA, a ocupar lugares intermedios al final del periodo.

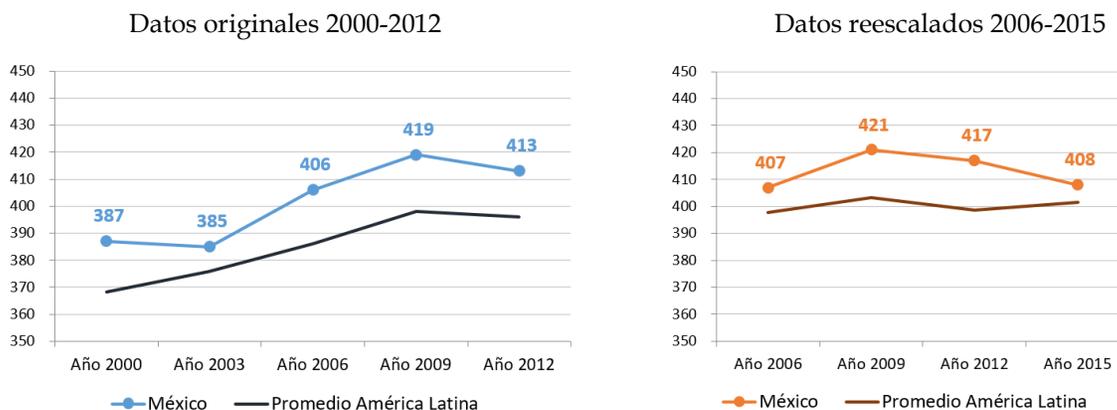
Gráfico 65. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

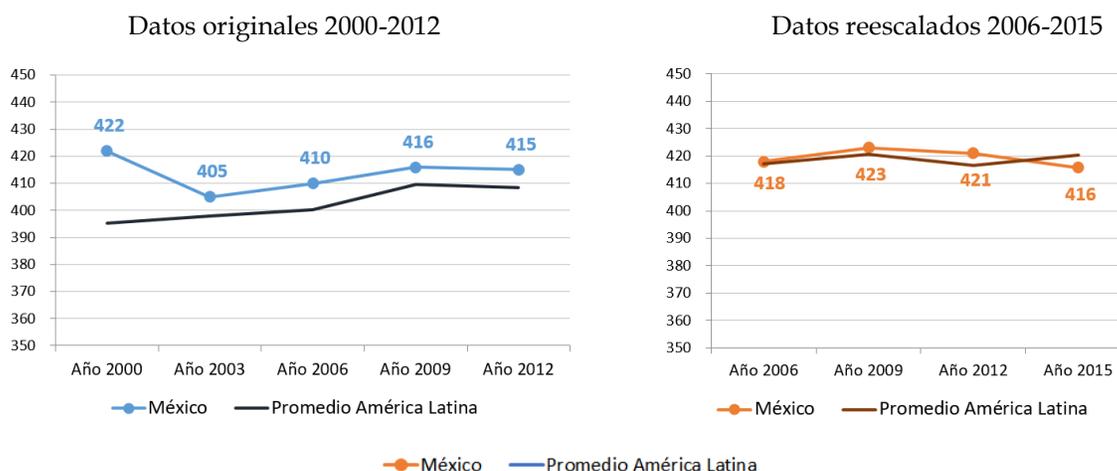
Al observar los cambios trienales, se complejiza la interpretación, ya que el país evidencia una caída generalizada en los desempeños 2000 - 2003 de 13 puntos promedio, con diferencias estadísticamente significativas en las tres áreas según los métodos de comparación utilizados en esa segunda aplicación. En las aplicaciones 2006 y 2009 el país muestra una mejora intensa y sostenida, con incrementos estadísticamente significativos en lectura (+26 puntos) y matemática (+33 puntos). Desde 2009 en adelante los resultados se estancan, con una leve caída.

Gráfico 66. Puntaje promedio en matemática, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Gráfico 67. Puntaje promedio en ciencias, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015.

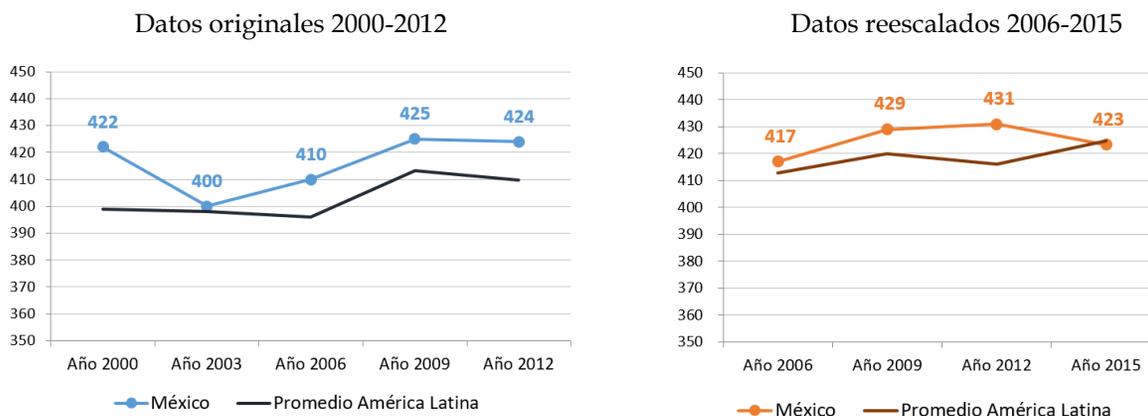


Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Este estancamiento evidenciado en los últimos años se corrobora al considerar los resultados reescalados con el cambio metodológico implementado en el último ciclo de PISA, disponibles desde 2006 en adelante³⁹. Con algunas diferencias entre áreas, los resultados reescalados del período muestran, las mismas tendencias: mejora en 2006 - 2009 de 10 puntos en promedio (estadísticamente significativas para lectura y matemática), y disminución de 9 puntos promedio 2009-2015. Esta reducción, aunque es mayor a la que perfilaban los resultados originales, no es estadísticamente significativa.

³⁹ Para mayor detalle, ver el capítulo 2 del documento.

Gráfico 68. Puntaje promedio en lectura, resultados originales y reescalados. México y promedio de países seleccionados de América Latina. 2000-2015.



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

En lectura es donde se intensifica la caída de los desempeños entre 2000 y 2003, donde el país retrocede 17 puntos (de 422 a 407 puntos). Estos resultados bajos se revierten en los años siguientes con mejoras en los desempeños (estadísticamente significativas entre 2003, 2006 y 2009). Ya para el 2009 México alcanza 425 puntos, niveles similares al 2000, y los resultados se mantienen estables en las aplicaciones siguientes.

La OCDE señala, en sus notas destacadas del país, que el rendimiento promedio de los jóvenes mexicanos de 15 años no varió significativamente en el área de ciencias desde el 2006. Y que, en promedio, el rendimiento en matemáticas mejoró en 5 puntos por cada tres años entre el 2003 y el 2015 (OCDE, 2016e). Lo que llama la atención de esta última afirmación es que la mirada global del período esconde las intensas oscilaciones en el área: México muestra, según los resultados comparables de la metodología 2015, una mejora entre 2006 y 2009 de 14 puntos (de 407 a 421 puntos), pero en 2015 el país obtiene 408 puntos, un deterioro en los desempeños estadísticamente significativo de 13 puntos.

En el informe nacional de los resultados de PISA de México se analizan estas tendencias, enfocándose en la evolución global de los puntajes para afirmar que los resultados del país muestran un comportamiento estable (INEE, 2016). No esgrimen hipótesis específicas que permitan explicar las oscilaciones en los resultados en ciertos períodos, especialmente cuando éstas toman forma de mejoras o retrocesos estadísticamente significativos, aunque resaltan ciertas debilidades en la comparación en el tiempo, asociados principalmente a los cambios metodológicos en el 2015.

Un elemento interesante a destacar en el informe es el análisis de la cobertura. Según las estadísticas nacionales, el porcentaje de población de 15 años que asiste a la escuela se incrementó sostenidamente en el período, del 57% en 2000 a 65% en 2006 y 78% en 2015. En este contexto, se resalta como logro que los resultados de los mexicanos en PISA se hayan mantenido casi sin cambio desde el año 2000, en el marco de una incorporación progresiva de estudiantes de esta edad históricamente excluidos de la educación secundaria y media (INEE, 2016).

Las evaluaciones nacionales de México permiten establecer una instancia de comparación con estos diagnósticos. El **recuadro 7** resume las principales características del sistema nacional de evaluación del país.

Recuadro 7. Las evaluaciones nacionales de la calidad educativa en México

El país desarrolló la Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE), que fue aplicada hasta el año 2014, último año de implementación. Actualmente, la política de evaluación de desempeños se encuentra en etapa de revisión. Las pruebas ENLACE se aplicaron en forma anual desde el año 2008 en el último año de la enseñanza media superior (12° año de estudios), desde el año 2006 hasta el 2013 en los últimos cuatro años de la educación básica (3° a 6° grado) y en el último año la educación secundaria. En el 2009 se incorporaron al ENLACE los restantes grados de la educación secundaria.

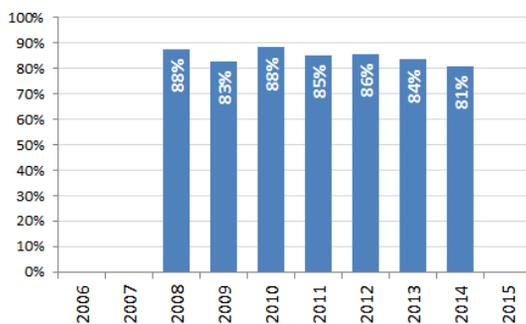
Las evaluaciones ENLACE son de frecuencia anual y de carácter censal, y se orientan a evaluar el desarrollo de competencias. Su aplicación es uniforme, lo que permite obtener resultados individuales. Se centran en dos campos disciplinares básicos, basados en los contenidos de los planes y programas de estudio vigentes: Comunicación (Comprensión Lectora) y matemáticas, y en la educación básica a partir de 2008 incluye una tercera asignatura rotativa. Los resultados se expresan en términos del porcentaje de estudiantes en cada uno de los siguientes cuatro niveles de logro: insuficiente, elemental, bueno y excelente.

También se incluyó la publicación de los resultados expresados en puntajes, con una escala estandarizada de media 500 y desvío estándar de 100. La difusión de resultados del ENLACE 2014 se realizó sólo en términos de niveles de logro.

Es posible comparar los resultados de la prueba ENLACE para el 3er año de la secundaria con las tendencias de PISA en términos de desempeños en lectura y matemática, al menos para el periodo 2008-2014, a los fines de identificar si existe confluencia de diagnósticos. La edad teórica para este grado son los 14 años de edad, lo que establece cierta equivalencia de comparación con la población objetivo de PISA, considerando los altos niveles de sobriedad en el país.

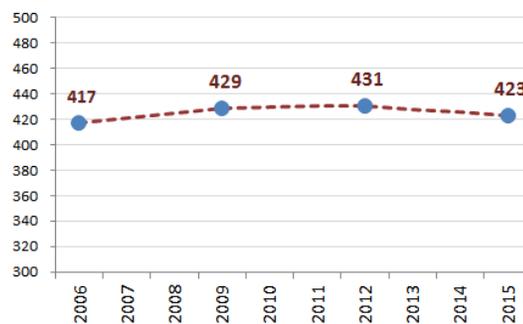
Gráfico 69. Resultados en lectura en pruebas nacionales e internacionales. Nivel secundario. México. 2006-2015

Evaluación Nacional - ENLACE 2008-2014
3er año secundaria



Unidad de medida: Porcentaje de estudiantes con logro elemental, bueno y excelente

PISA 2006 - 2015
Resultados reescalados - 15 años



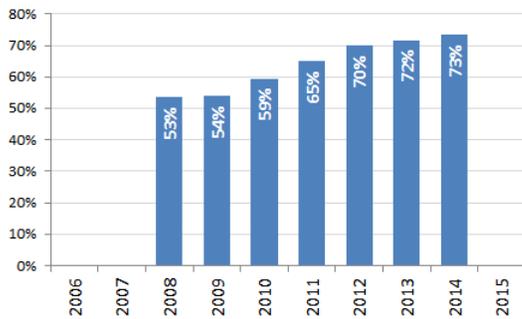
Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados ENLACE 2008-2014 y OCDE-PISA 2006-2015.

Para este año de estudio, los resultados de ENLACE muestran tendencias contrapuestas entre áreas. En lectura, el 87,6% de los evaluados alcanzaba niveles de logro elemental, bueno y excelente en el 2008, y este porcentaje se reducía al 80,9% en el 2014, con una tendencia a la baja localizada entre los años 2010 y 2014. Esta reducción está principalmente explicada por una caída en el porcentaje de estudiantes con logros excelentes (de 6,7% a 4,6%) y buenos (del 45,5% al 40,1%).

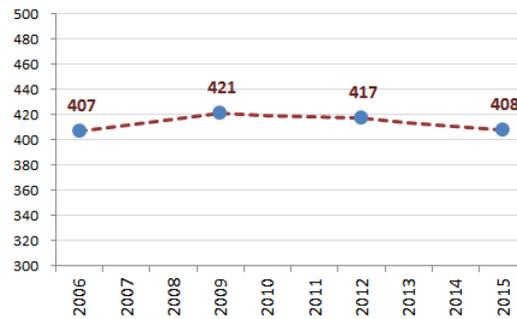
Gráfico 70. Resultados en matemática en pruebas nacionales e internacionales. Nivel secundario. México. 2006-2015

Evaluación Nacional - ENLACE 2008-2014
3er año secundaria



Unidad de medida: Porcentaje de estudiantes con logro elemental, bueno y excelente

PISA 2006 - 2015
Resultados reescalados - 15 años



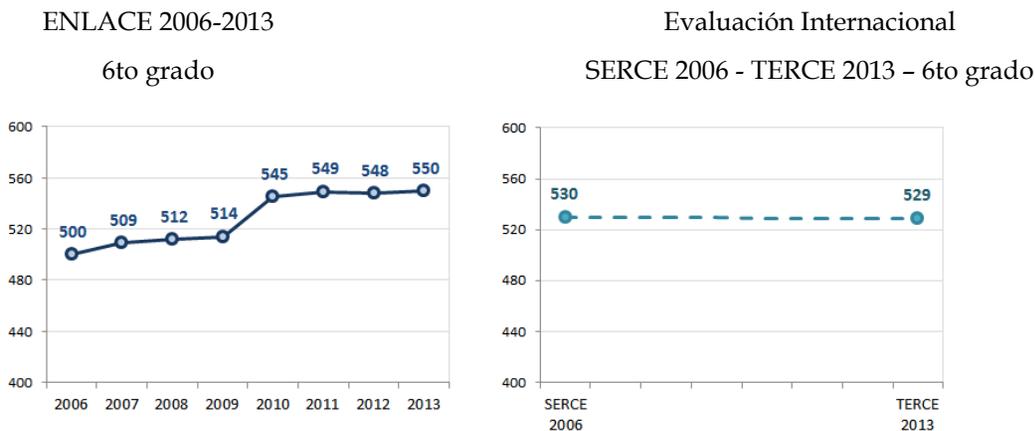
Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados ENLACE 2008-2014 y OCDE-PISA 2006-2015.

En cambio, la tendencia de los resultados en matemática muestra una marcada mejoría en el país, donde el porcentaje de los evaluados que alcanzaba niveles de logro elemental, bueno y excelente ascendía del 53% en 2006 al 73% en 2014. Esta mejora se intensifica entre 2009 y 2012, donde los resultados mejoran 16 puntos porcentuales en tres años. En matemática, el porcentaje de estudiantes con logros excelentes se incrementa del 3,4% al 19,4% en estos años.

Un aspecto interesante para incorporar en el diálogo entre diferentes resultados es la tendencia de los desempeños de los estudiantes del nivel primario. Para ello, se analizan las tendencias de los desempeños de los estudiantes de 6to grado en las pruebas ENLACE entre 2006 y 2013.

Gráfico 71. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. México. 2006-2013



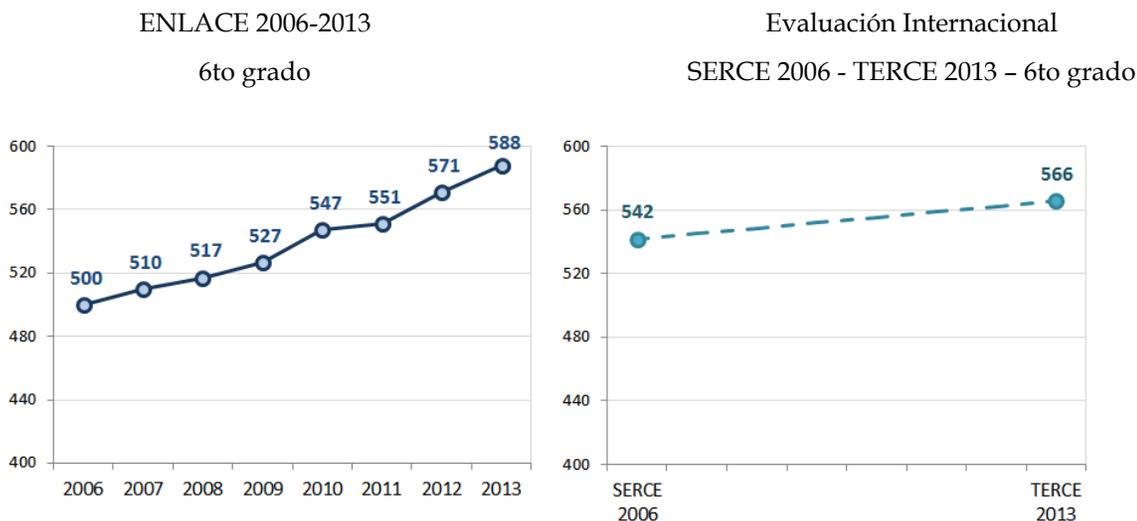
Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar =100

Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados ENLACE 2006-2013 y SERCE - TERCE.

El primer aspecto destacable es el marcado incremento de los resultados en el período. En lectura, los desempeños se incrementan de 500 a 550 puntos (lo que representa medio desvío estándar en la escala). Lo llamativo es que gran parte de ese crecimiento se concentra entre 2009 y 2010, donde el puntaje crece 31 puntos. En el resto de los años, se observa estabilidad o crecimiento leve.

Gráfico 72. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. México. 2006-2013



Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar =100

Unidad de medida: Puntaje promedio estandarizado con media = 500 y desvío estándar =100

Fuente: CIPPEC, sobre la base de resultados ENLACE 2006-2013 y SERCE - TERCE.

En matemática, la mejora del desempeño es aún más intensa, con un incremento de 88 puntos en los 8 años representados en el período. En este caso, la mejora ha sido sostenida año a año, con diferencias interanuales que oscilan entre los 3 y 20 puntos.

Las pruebas del LLECE-UNESCO son la referencia internacional para analizar estas tendencias en el nivel primario en la comparación con otros países de América Latina. Los resultados de aprendizaje de nivel primario (6to grado) del año 2013 posicionan al país en lugares altos en el contexto de la región. México ocupa el 4to lugar en lectura y el 3ro en matemática, en el conjunto de 14 países evaluados. La mirada comparada SERCE (2006) y TERCE (2013) permite identificar las tendencias de cambio en los desempeños por área. Se destaca el incremento de matemática entre 2006 y 2013 en magnitudes superiores al conjunto de países: pasó de 542 a 566 puntos (24 puntos), por encima de la mejora de 17 puntos en el conjunto de países. En lectura el país no evidencia cambios significativos: los puntajes se redujeron levemente entre ambas aplicaciones, de 530 a 529 puntos.

Es posible identificar algunos elementos de diagnóstico de las distintas evaluaciones que permiten reconocer tendencias convergentes, y algunas diferencias y particularidades. El escenario general tiende a mostrar al país con desempeños medios a altos en relación a los 7 países del estudio. Es destacable a su vez que es el país con mayor porcentaje de población fuera de la escuela.

La mayoría de las fuentes analizadas tienden a presentar un escenario de estabilidad de resultados en el nivel secundario, y ciertas mejoras en el primario. Este “estancamiento” de los desempeños del nivel secundario, leído en el contexto de un incremento de la cobertura, señala el logro del país de lograr sostener los resultados de aprendizaje en un contexto de incorporación de población históricamente excluida de esta etapa, aspecto resaltado en el informe nacional. Los resultados de EXCALE de 3° año del secundario parecen corroborar este escenario de estabilidad.

Otro aspecto destacable es la diferencia observada entre áreas: tanto las evaluaciones ENLACE de primaria y secundaria, como la comparación SERCE-TERCE tienden a mostrar resultados estables en el área de lectura, y de mejora sostenida en matemática. PISA es la única evaluación que no expresa esas diferencias.

Resumen de datos educativos y sociales destacados de México

Cobertura

1. Entre los países estudiados, los niveles históricos de cobertura de México eran altos para el nivel primario: ya en 1996, la tasa neta de cobertura en este nivel era del 97% (SEDLAC).
2. En el nivel inicial la matrícula se mantiene estable, incrementándose levemente la tasa bruta de matrícula de 3 a 5 años entre 2008 y 2011 (de 70% a 72%), por debajo del promedio de los siete países estudiados (76%) (SEDLAC).
3. Sin embargo, la mayor deuda se refleja en el acceso a la educación secundaria: en el año 2014, la tasa neta de escolarización de nivel secundario es del 78,3%, la más baja de los siete países estudiados. Cabe destacar el esfuerzo de expansión del nivel, que en la última década creció casi 15 puntos, el primer incremento más importante del periodo (SITEAL).

Trayectorias educativas

1. La trayectoria de los alumnos tiene dos rasgos salientes: baja repitencia y alta exclusión en la educación media superior.
2. En cuanto a los niveles de repitencia en el nivel primario, al año 2015 sus valores son los más bajos (0,8%), en relación al promedio de los 7 países. A su vez, el indicador muestra una tendencia a la mejora durante los últimos diez años. Por otro lado, el porcentaje de estudiantes de 15 años que repitieron al menos una vez se encuentra en el lugar más bajo en relación a los 7 países (15,8%).

Equidad

1. En los resultados de equidad, al año 2015, México es el país con más bajas brechas de resultados entre los cuartiles de mayor y menor nivel socioeconómico (62 puntos en 2015, lectura y matemática en conjunto, frente a un promedio de 84 puntos de diferencia en el conjunto de siete países). Adicionalmente, México logró disminuir marcadamente la brecha entre los alumnos de mayor y menor nivel socioeconómico entre 2000 y 2012, siendo la brecha a inicios del periodo de 99 puntos. Es importante considerar que hay una relación entre estas brechas reducidas y el amplio porcentaje de población no evaluada.
2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que México es el cuarto país con mayores niveles de segregación educativa en función del nivel socioeconómico, dentro del universo de países estudiados. Esta situación se redujo levemente respecto de los niveles de segregación del año 2000.

Contexto

1. Durante la primera mitad de la década del 2000 se perfila un mejoramiento de las condiciones de vida de la población en México. Según las estadísticas oficiales del país, el porcentaje de población bajo la línea de pobreza descendió del 31,8% en 2000 al 20,9% en 2006. Sin embargo, el indicador se estanca en la segunda mitad de la década, incrementándose hacia el año 2014, donde alcanza el 29,1%. Un indicador comparable en la región, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, muestra la siguiente tendencia: mejora en la primera mitad de la década (del 59,2% al 55,0% entre 2000 y 2006), posterior estancamiento entre 2008 y 2012 (55,6% y 52,4%) y una leve disminución en 2014 (48,2%). A su vez, el indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) muestra una tendencia similar: una mejoría en las condiciones de vida de la población de 15 años entre 2000 y 2006, y un estancamiento en la segunda mitad.
2. Esta evolución de la pobreza ocurre en un contexto de mejora de los indicadores macroeconómicos: desde el inicio de la década del 2000 al 2014 el PBI per cápita se incrementó en el país en casi un 46% en valores constantes a 2008 (IMF-WEO). En este periodo, la evolución anual del PBI se muestra oscilante, con algunos años de mayor expansión (5,3% en el 2000, 4,3% en 2004, 5,0% en 2006, 5,1% en 2010), y otros años de estancamiento e incluso de retroceso (-0,6% en 2001, -4,7% en 2009, 1,4% en 2003).
3. En estos años (2000-2015) ha mejorado la distribución social de la riqueza, medida a través del coeficiente de Gini, con un descenso del 0,54 al 0,49. Sin embargo, esta mejora también se

concentra en la primera mitad de la década: para el año 2006, el coeficiente de Gini alcanzaba el valor de 0,50 (SEDLAC), mientras que entre los años 2012 y 2015 se mantuvo estable.

4. México manifiesta niveles muy bajos de empleo formal (el 62% de los trabajadores no posee cobertura de pensión, contra el 33% promedio de los siete países), con niveles bajos de desempleo (4,3% en el país – el valor más bajo de los siete países estudiados – y 6,4% en el promedio). Pero también tiene niveles muy bajos de cobertura de salud (sólo el 47% de la población ocupada urbana posee cobertura de salud, uno de los valores más bajos del conjunto, similar a Colombia, y muy por debajo del promedio de los 7 países, de 70,4%).

Inversión educativa

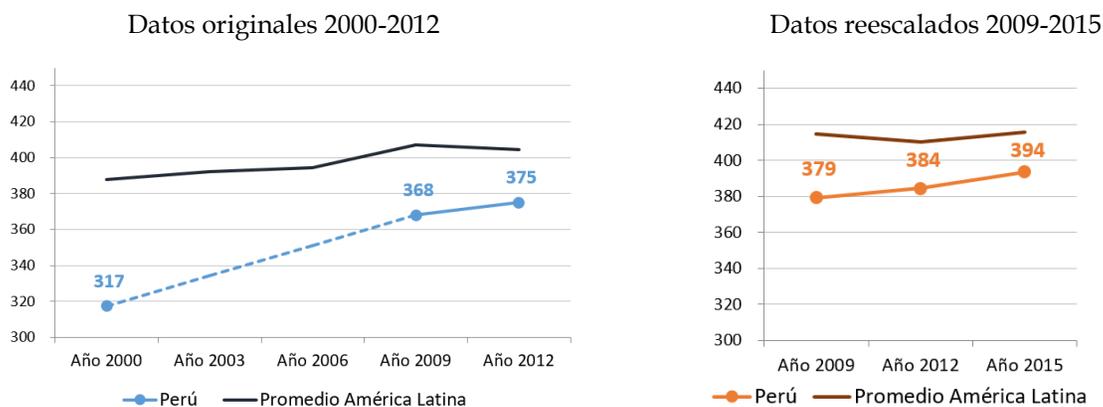
1. México es uno de los países que aumentó su inversión en el sistema educativo en la última década. Medido como porcentaje del PBI, el gasto en educación creció del 4,1% en el 2000 al 5,1% en 2012. Este crecimiento es, en puntos porcentuales, similar al promedio del conjunto de países de América Latina en el periodo estudiado. Gracias a este crecimiento, al año 2012 alcanza niveles de inversión (medidos como % del PBI) altos en la región (el promedio de 16 países de la región es 5,0% en valores *circa* 2013)
2. Asimismo, el país destina, dentro del gasto público total, el 19,0% a educación (año 2013). En el promedio de 16 países este porcentaje era del 14,3% en valores *circa* 2012. Esta porción de los recursos públicos destinados a educación es muy similar en relación al 2000, año en el que se destinó a educación el 19,7% del gasto público.

4.6. Perú

Resultados comparados de las evaluaciones de la calidad educativa

En las pruebas PISA, Perú participó en forma discontinua en las primeras aplicaciones. Fue evaluado en el año 2000, y no participó en las sucesivas aplicaciones hasta el año 2009. De allí en adelante fue evaluado en todas las aplicaciones subsiguientes. Los resultados de las pruebas en el país perfilan un escenario global de marcado incremento en todas las áreas evaluadas, con variaciones que van de los 317 puntos en el 2000 a los 375 en el 2012 según los resultados originales de cada implementación (promedio 3 áreas).

Gráfico 73. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015



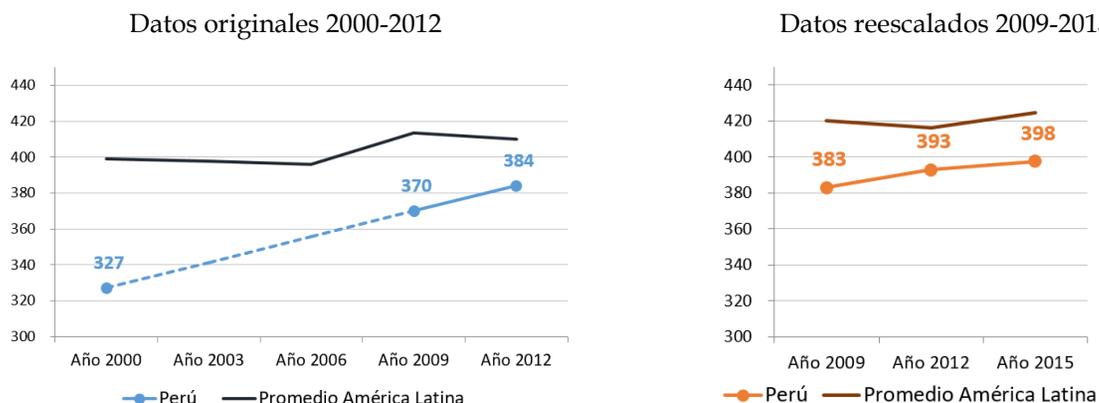
Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

La mirada sobre los resultados comparables, reescalados con el cambio metodológico implementado en el último ciclo de PISA, permite precisar la mejora en los desempeños⁴⁰. Con algunas diferencias entre áreas, los resultados reescalados del período muestran tendencias crecientes entre 2009 y 2015. El incremento promedio de 26 puntos entre 2009 y 2015 que se identifica en los resultados originales, se reduce a 14 puntos con los reescalados (que son estrictamente comparables). Estas diferencias, aunque de menor magnitud, continúan siendo estadísticamente significativas en casi todos los casos.

Si bien Perú aún permanece en el conjunto de países con más bajos resultados en PISA 2015, es muy destacable la sustantiva mejora, que arranca desde un punto de partida muy bajo en el año 2000. Tal como se destaca en el informe nacional de PISA (Ministerio de Educación, 2017), Perú no solamente es el país que muestra mayor crecimiento en el periodo global de aplicaciones de PISA, sino también es el país que más creció en la región entre las aplicaciones 2012 y 2015 (considerando los resultados comparables).

⁴⁰ Para mayor detalle, ver el capítulo 2 del documento.

Gráfico 74. Puntaje promedio en lectura, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015

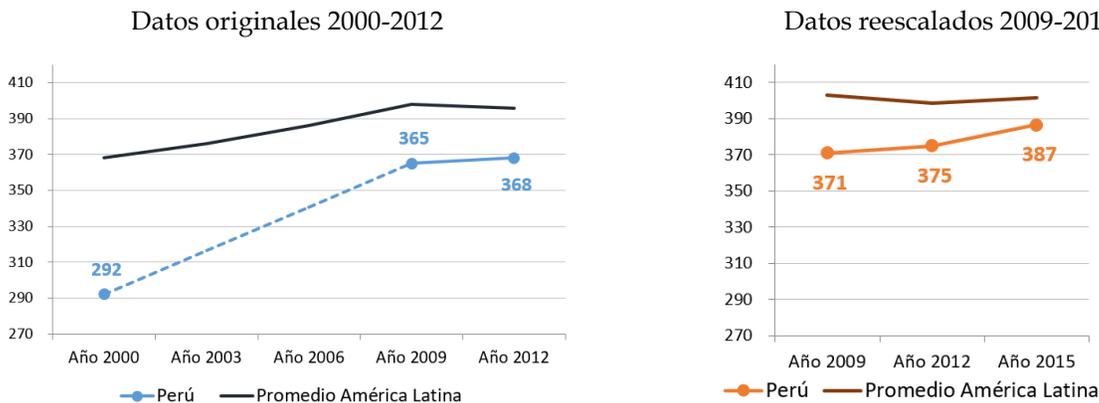


Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

En lectura es donde el país muestra una mejora continua en todas las aplicaciones de PISA. Desde un punto de partida muy bajo (327 puntos en el año 2000, el resultado más bajo de todos los países evaluados ese año), los resultados se incrementaron sostenidamente desde la siguiente aplicación (año 2009, con 383 puntos calculados con la metodología 2015), con mejoras de 16 puntos al 2015 (estadísticamente significativas).

El informe nacional destaca esos logros (pero refiriéndose a los resultados originales), y señala que Perú es el único país de la región que muestra una mejora sostenida desde PISA 2000, inclusive en 2009 - 2012, cuando el resto de los países de la región disminuyen en sus desempeños (Ministerio de Educación, 2017).

Gráfico 75. Puntaje promedio en matemática, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

Gráfico 76. Puntaje promedio en ciencias, resultados originales y reescalados. Perú y promedio de países seleccionados de América Latina. 2000-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2000-2015.

En matemática y ciencias se observan las mismas tendencias crecientes. Los resultados del año 2000 posicionaban al país en el lugar más bajo, aunque no son estrictamente comparables con las evaluaciones siguientes, permiten identificar el esfuerzo realizado por el país para mejorar sus desempeños. En ambas áreas se identifica una mejora sustantiva en el 2009. Utilizando la misma metodología de PISA 2015, Perú hubiese obtenido 375 puntos en matemática en 2012, lo que implicaría identificar con datos comparables una mejora estadísticamente significativa de 16 puntos en las dos últimas aplicaciones (menor a los 21 puntos que arrojan los resultados originales). En forma similar, en el caso de ciencias los resultados reescalados del 2012 son de 385 puntos, lo que arroja una mejora de 13 puntos (en lugar de los 27 puntos de diferencia en los resultados originales), crecimiento también estadísticamente significativo.

Otro aspecto que se destaca en el informe nacional (Ministerio de Educación, 2017) es el análisis de la cobertura: según las estadísticas informadas a PISA por el país, el porcentaje de población de 15 años que asiste a la escuela se incrementó marcadamente entre los ciclos de PISA, del 76,4% en 2000 al 90,3% en 2015. Adicionalmente, se reduce el porcentaje de la muestra con atraso escolar, del 47,1% al 24,8%. En otras palabras, se resalta como significativo logro del país el hecho de que los resultados se hayan incrementado sostenidamente en un contexto de ampliación del acceso a la educación secundaria, con una incorporación progresiva de estudiantes históricamente excluidos de la educación secundaria, y con una marcada mejora en los indicadores de trayectorias.

Al comparar con otras evaluaciones, cabe señalar las limitaciones existentes, dado que Perú aplica anualmente la Evaluación Censal de Estudiantes (ECE) desde el año 2007 en forma censal en el 2do grado de la primaria. Existe una significativa distancia entre la población objetivo de esta evaluación y la de PISA, que implican mucho resguardo en la comparación de tendencias. En los últimos años incorporó una evaluación a una muestra de 4to grado, innovadora al incorporar una evaluación del castellano como segunda lengua para diferentes comunidades indígenas. Más recientemente incorporó una evaluación al 2do año del secundario, de la que no se disponen más que resultados del 2015, lo que impide una mirada de tendencias.

Recuadro 8. Las evaluaciones nacionales de la calidad educativa en Perú

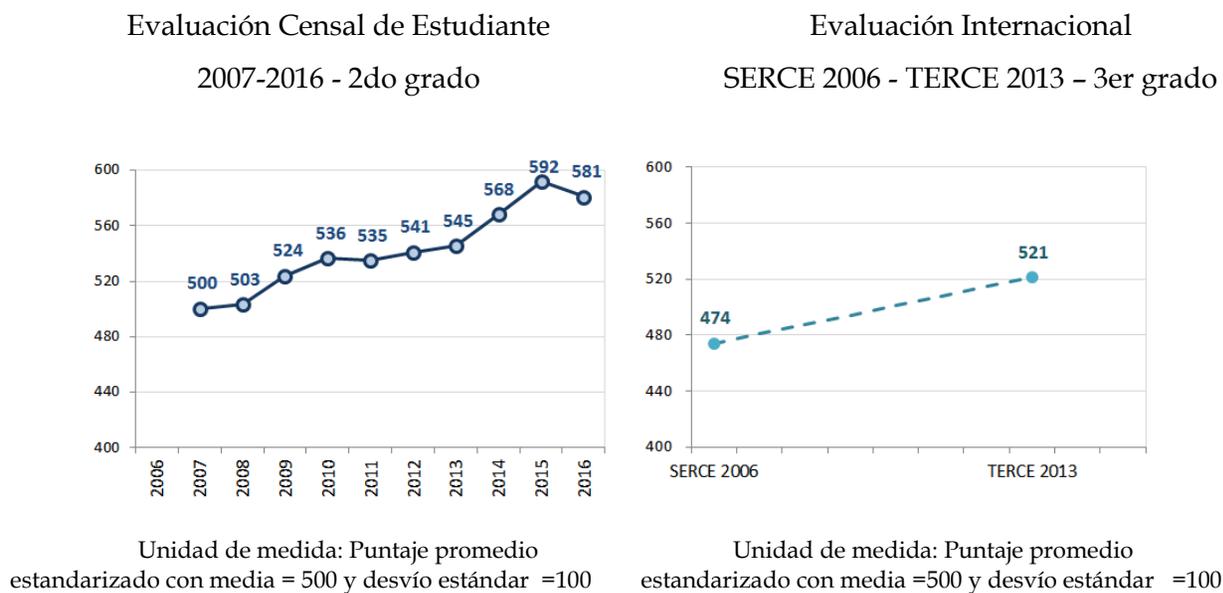
Desde 2007, Perú aplica anualmente la Evaluación Censal de Estudiantes (ECE) en el 2do grado de primaria. El organismo responsable es la Unidad de Medición de la Calidad Educativa del Ministerio de Educación de la Nación.

La ECE es una evaluación alineada al currículum, del tipo criterial, que evalúa las áreas de lengua y matemática. Se aplica de forma censal a los alumnos de 2do año del nivel primario, y de forma muestral en el 4to año, a estudiantes cuya lengua materna originaria es distinta al castellano y asisten a una escuela que desarrolla Educación Intercultural Bilingüe. En 2016 se incorporó como población evaluada a los estudiantes de 4to año del nivel primario (de Educación Básica Regular) y de 2do año de nivel secundario.

En cuanto a los resultados, estos se representan en una escala con media 500 y desvío de 100, y también se expresan en niveles de logro: en inicio, en proceso y satisfactorio.

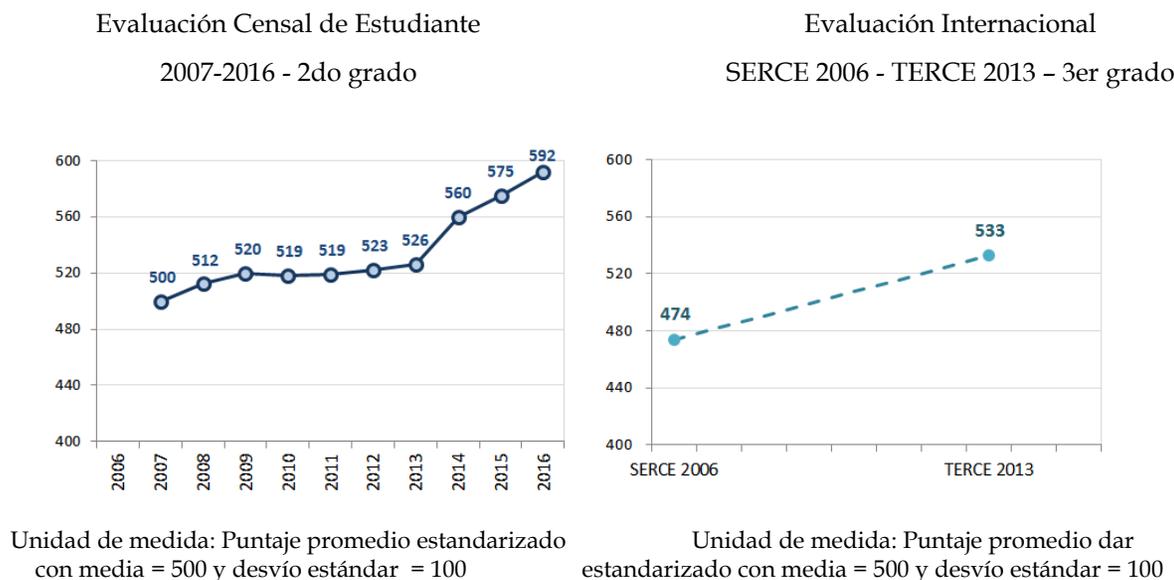
Su antecedente es la Evaluación Nacional del Rendimiento Estudiantil (ENRE), que fue aplicada en 1996, 1998, 2001 y 2014 de forma muestral en las áreas de lengua, matemática y ciencias sociales.

Gráfico 77. Resultados en lectura en pruebas nacionales e internacionales. Nivel primario. Perú. 2007-2016



Fuente: CIPPEC, sobre la base de resultados ECE 2007-2016 y SERCE - TERCE.

Gráfico 78. Resultados en matemática en pruebas nacionales e internacionales. Nivel primario. Perú. 2007-2016



Fuente: CIPPEC, sobre la base de resultados ECE 2007-2016 y SERCE - TERCE.

Lo que muestra la ECE de 2do grado es una mejora muy marcada de los desempeños entre 2007 y 2015. En lectura, el puntaje se incrementa de 500 a 592 puntos, y en matemática de 500 a 575. En una escala que posee un desvío estándar de 100 puntos, estas diferencias son muy significativas. Cabe resaltar que si bien la tendencia es creciente en general, en ambas áreas se observa una importante mejora en los resultados entre los años 2013 y 2014.

Las pruebas TERCE del LLECE-UNESCO confirman estas tendencias de mejora. Los resultados de aprendizaje de nivel primario (3er grado) del año 2013 posicionan al país en lugares medio/altos en el contexto de la región. En matemática, Perú ocupó con 533 puntos el 7mo lugar entre 15 países, por encima del promedio de países (522). En lectura, los desempeños fueron más destacados, alcanzando el 4to lugar en el conjunto de países evaluados. Lo que resalta del caso peruano es la mejora en los desempeños respecto del SERCE: se destaca el país como el que más creció en matemática entre 2006 y 2013, pasando de 474 a 533 puntos (59 puntos de diferencia, en un contexto donde el promedio de mejora fue de 31). En lectura también muestra una mejora por encima del promedio (47 puntos), sólo superada por tres países.

El escenario general de Perú tiende a mostrar al país con desempeños medios a bajos en relación a los siete países del estudio, aunque existe una coincidencia en todas las fuentes analizadas de presentar un escenario de mejora a ritmos muy superiores a la media de países. Si bien el país no dispone de una evaluación que permita contrastar las tendencias en poblaciones de edad similar, la mirada sobre el nivel primario contribuye a identificar tendencias confluyentes. Otro aspecto destacable es que el país mejora en calidad, cobertura y trayectorias al mismo tiempo, lo que posiciona a Perú en un lugar muy destacado respecto de otros países de la región.

Resumen de datos educativos y sociales destacados de Perú

Cobertura

1. Perú es un caso destacado por sus avances en la cobertura escolar. Parte de puntos más bajos que otros países estudiados: para el año 2000, la tasa neta de escolarización secundaria se situaba en el 68,8% (ESCALE), una de las más bajas de los siete países estudiados después de México y Colombia (SITEAL- SEDLAC). Incluso la tasa de analfabetismo alcanzaba, para 1980, casi a una quinta parte de la población de 15 años y más (18,1%).
2. Con este punto de partida tan bajo, Perú expande intensamente el acceso a la educación secundaria, con un crecimiento que al año 2014 alcanza 14,1 puntos porcentuales, situando a la tasa neta de escolarización secundaria en un 82,9%, valor sólo por debajo de Argentina y Chile, y encima del promedio de los 7 países de la región de 81,5% (Circa 2014). Este importante crecimiento beneficia principalmente a las clases bajas: la tasa neta de escolarización del quintil más pobre aumenta del 35% al 78% en el periodo (SEDLAC).
3. También logró grandes avances en el nivel inicial a lo largo del período estudiado: entre 2000 y 2014 la tasa de crecimiento de la matrícula inscrita en el nivel fue del 28%, incorporando a más de 500 mil estudiantes nuevos. Este crecimiento es superior al total de América Latina y el Caribe (24,0%), e implica incorporar al 79% de la población de 3 a 5 años al sistema educativo.

Trayectorias de los alumnos

1. Las tasas de repitencia en el nivel primario en el año 2014 era de 3,2%, encontrándose por debajo de países como Brasil, Uruguay y Chile. Se destaca la reducción de la repitencia en el este nivel, ya que en el 2002 la repitencia era del 10,7%.
2. La población de 15 años que asistía a la escuela presentaba un 25,6% de estudiantes que repitieron al menos una vez. Este porcentaje se encontraba por debajo del promedio regional (30%), según lo informado en el cuestionario complementario de PISA 2015.

Equidad

1. En los resultados de equidad, se observan indicadores preocupantes: es el país de América Latina con más altas brechas de resultados entre cuartil de mayor y menor nivel socioeconómico (106 puntos en 2015, lectura y matemática en conjunto, frente a un promedio de 86 puntos de diferencia en el conjunto de siete países). A pesar de producirse una leve disminución de la brecha entre 2012 y 2015 de 4 puntos, fue el único país de los siete donde se incrementó la brecha de resultados entre los cuartiles de menor y mayor nivel socioeconómico (de 100 a 106 puntos entre 2000 y 2015). Esto indica que el país logró mejorar los resultados de calidad pero a costa de aumentar su desigualdad.
2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que Perú es el país con mayores niveles de segregación educativa en función del nivel socioeconómico, dentro del universo de países estudiados.

Contexto

1. Durante la última década se perfila un marcado mejoramiento de las condiciones de vida de la población en Perú. Según las estadísticas oficiales del país, el porcentaje de población bajo la línea de pobreza descendió del 48,4% en 2000 al 21,7% en 2015. Un indicador comparable en la región, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, también muestra una mejora significativa en el periodo, del 48% al 38%. A su vez, el indicador de pobreza construido por PISA (índice del status económico, social y cultural, ESCS por sus siglas en inglés) perfila también una marcada mejora en las condiciones de vida de la población de 15 años entre 2000 y 2015. Según esta fuente, la población adolescente de Perú evaluada en PISA se encuentra en el grupo de países de mayor pobreza entre los siete evaluados, junto con México.
2. Esta evolución de la pobreza es consistente con el comportamiento de los indicadores macroeconómicos: desde el inicio de la década del 2000 hasta el 2014 el PBI se incrementó constantemente en el país, mostrando un crecimiento anual promedio de 5,3%, en valores constantes a 1994 (IMF-WEO).
3. En estos años de crecimiento económico ha mejorado en forma marcada la distribución social de la riqueza, medida a través del coeficiente de Gini, con un importante descenso del 0,51 al 0,44 (es el tercer país con mayor mejora en la distribución del ingreso entre 2000 y 2014, después de Argentina y Brasil). Al año 2015, Perú manifiesta niveles intermedios de desigualdad en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados es de 0,47 (SEDLAC).
4. Si bien los niveles de empleo formal en Perú son altos, ha manifestado una mejora entre los años 2012 y 2014: en el periodo se redujo del 69% al 52% la proporción de trabajadores que no poseen cobertura de pensión (el promedio de los siete países es del 46%), se alcanzaron niveles medio de desempleo (6% en el país y 6,6% en el promedio), y bajos niveles de cobertura de salud (el 63,2% de la población ocupada urbana posee cobertura de salud, cuyo promedio en el conjunto de países es de 70%).

Inversión educativa

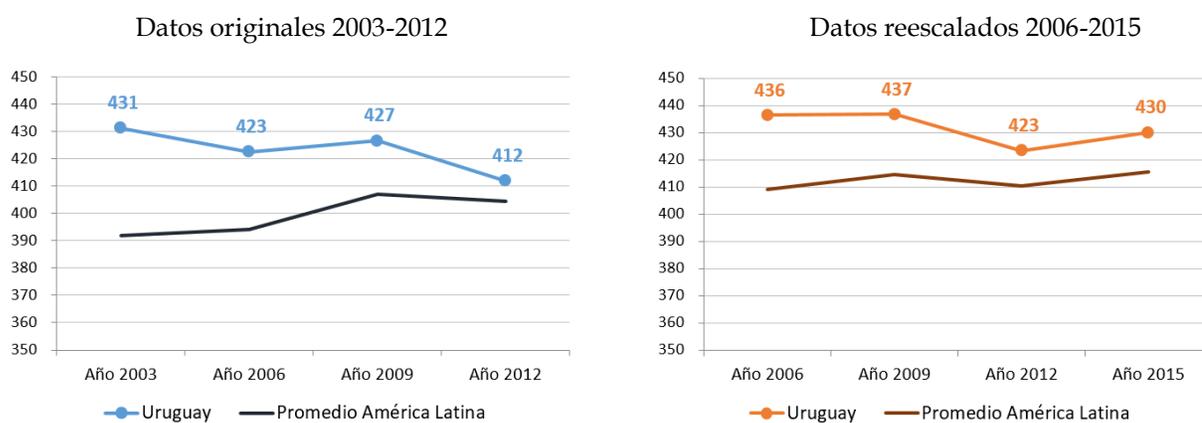
1. El gasto educativo en Perú, como porcentaje del PBI, se mantuvo contante considerando el periodo completo 2000-2014, se incrementó levemente del 3,2% al 3,7%. Sin embargo, en este lapso de tiempo muestra un periodo de decrecimiento (alcanza sus valores más bajos en 2007 con un gasto educativo del 2,6% del PBI), y crece más intensamente en los últimos años, pasando del 2,7% (2011) al 3,7% (2014)
2. Al año 2014, ocupa lugares medio - bajos en términos de inversión educativa, considerando el conjunto de países de América Latina (es el 6to país con valores más bajos en gasto educativo como % del PBI). El promedio de 16 países de la región es 5,0% en valores *circa* 2013.
3. Asimismo, es uno de los países que destina, dentro del gasto público total, porcentajes medios a educación. Al año 2013, el 15,6% de los recursos públicos se destinaban al sistema educativo, mientras que en el promedio de los 16 países este porcentaje era del 14,3% en valores *circa* 2012. Esta porción de los recursos públicos destinados a educación se mantiene relativamente estable. Entre los años 2000 y 2012 el valor ha variado poco, del 13,9% (2011) al 16,0% (2000).

4.7. Uruguay

Resultados comparados de las evaluaciones de la calidad educativa

En las pruebas PISA, Uruguay participó en forma continuada desde el 2003 al 2015, lo que permite contar con una serie de tiempo amplia para interpretar los resultados. Un elemento desafiante para el análisis y la comprensión de las tendencias observadas es la diferencia de diagnóstico al comparar los resultados publicados originalmente por PISA y los reescalados con el cambio metodológico implementado en el último ciclo de PISA que permite la comparación correcta⁴¹.

Gráfico 79. Puntaje promedio en las 3 áreas evaluadas, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015



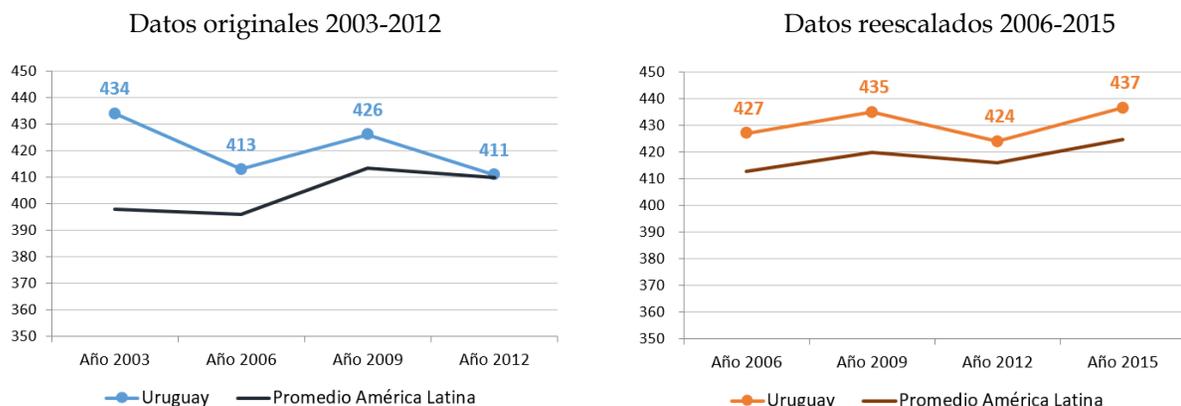
Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2003-2015.

La evolución con los resultados reescalados muestra entre 2006 y 2015 un escenario de estancamiento, e incluso de retroceso entre 2009 y 2012. En las aplicaciones 2012 y 2015 los desempeños de Uruguay mejoraron, 7 puntos en promedio, aunque estas diferencias no son estadísticamente significativas para ninguna de las tres áreas. Es importante resaltar en el caso de Uruguay que, con la metodología utilizada en las publicaciones anteriores, Uruguay obtuvo 412 puntos en PISA 2012, lo que arrojaría un incremento de 18 puntos, promedio de las tres áreas (mejora estadísticamente significativa en las tres) entre 2012 y 2015. Esta diferencia entre estimaciones se vincularía, como se analiza en el capítulo 2, a que Uruguay es uno de los países que muestra mayores porcentajes de pruebas con ítems incompletos al final (*non-reached items*).

La significativa diferencia entre los resultados reescalados (comparables) y los originales es lo que coloca al país en una situación confusa para perfilar la evolución reciente de los desempeños.

⁴¹ Para mayor detalle, ver el capítulo 2 del documento.

Gráfico 80. Puntaje promedio en lectura, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015



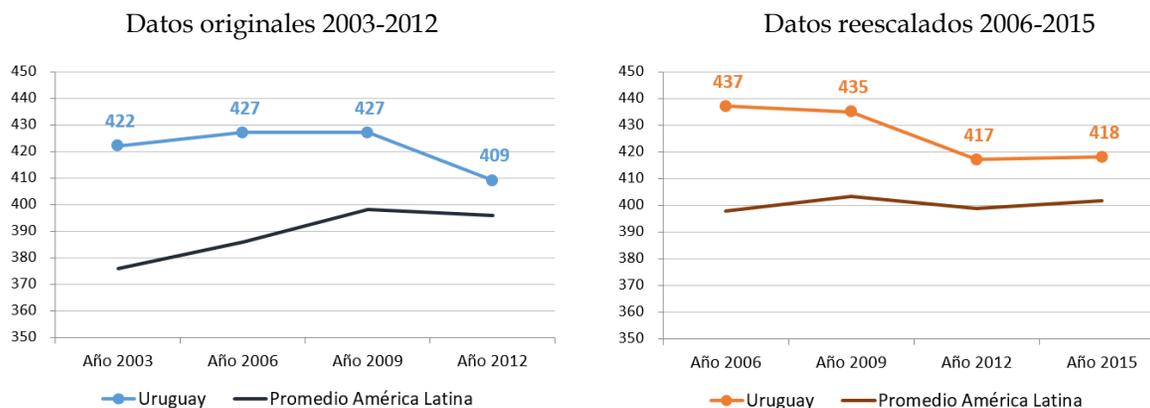
Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2003-2015.

La OCDE, en sus notas destacadas del país, resalta los resultados de lectura obtenidos en Uruguay, indicando que tanto el puntaje promedio como su ubicación entre los países latinoamericanos mejoraron entre 2009 y 2015, con una mejora promedio de 5 puntos cada tres años (OCDE 2016g), diferencia estadísticamente significativa. Asimismo, se resalta que Uruguay superó a México y a Costa Rica (cuyos puntajes medios eran, respectivamente, similares y superiores). Paradójicamente, esta mejora sustantiva queda puesta en cuestión con los reescalamientos de resultados que PISA realiza a partir de los cambios metodológicos implementados en 2015: la diferencia de puntajes se reduce a una estabilidad cuando se compara de manera estricta las pruebas de los distintos años.

En el documento de notas destacadas del país, la OCDE también introduce la hipótesis de que estos cambios de tendencias tendrían su principal causa en el cambio en la forma de considerar a los *non-reached items*⁴². Esta explicación no resulta del todo clara para establecer en qué medida la mejora en los puntajes se explica por un mejor rendimiento (mayor porcentaje de respuestas correctas) o por un cambio en la metodología.

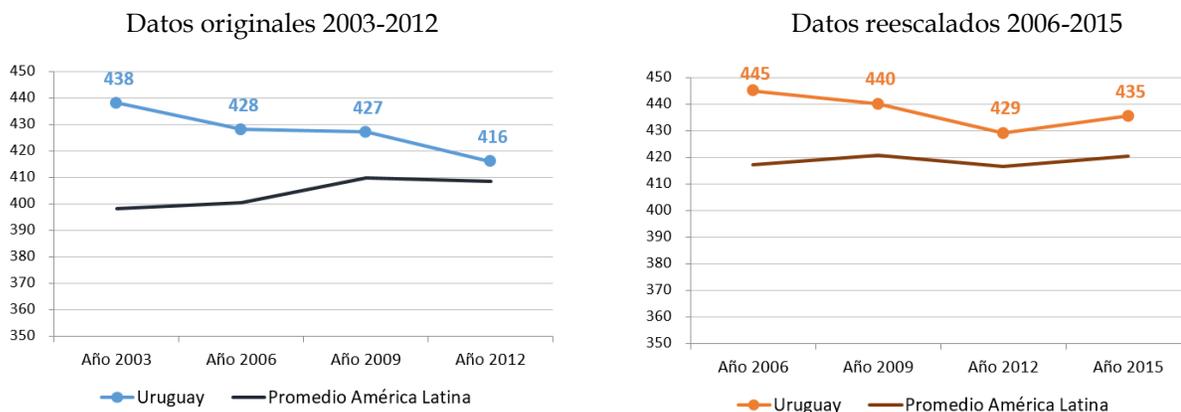
⁴² Para mayor detalle, ver el capítulo 2 del documento.

Gráfico 81. Puntaje promedio en matemática, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2003-2015.

Gráfico 82. Puntaje promedio en ciencias, resultados originales y reescalados. Uruguay y promedio de países seleccionados de América Latina. 2003-2015



Fuente: Fuente: CIPPEC, sobre la base de OCDE-PISA 2003-2015.

Por su parte, el informe nacional de PISA 2015 manifiesta que el país alcanzó los mejores puntajes promedio desde que participa en la evaluación en ciencias, con un puntaje promedio de 435 puntos. Estos resultados pondrían en evidencia una mejora generalizada con respecto al ciclo 2012 (ANEP, 2016a). Ahora bien, si se utiliza la metodología de escalamiento de PISA 2015, Uruguay hubiese obtenido 445 puntos en ciencias en 2006, y 440 en el 2009. Es decir, los resultados del 2015 apenas superarían a los del 2012 (429 puntos), dato que introduce interrogantes al avance del país en sus desempeños.

En las notas metodológicas sobre las estimaciones de los resultados PISA 2015 (ANEP, 2016b) el país hace mención al problema de los cambios mencionados, señalando que “el análisis de tendencias podría verse afectado por estas modificaciones metodológicas”. Allí quedan también expresadas las dificultades para adoptar uno u otro criterio en la caracterización de las tendencias en

el tiempo. El propio informe reconoce que “los países participantes no cuentan a la fecha con la información necesaria para replicar este ajuste con sus bases de datos nacionales. En consecuencia, los informes de los países, incluido el de Uruguay, trabajaron en base a la metodología regular. De hecho, tal como se expresó antes, el Informe Internacional de la OCDE también realiza sus análisis de tendencia con esta misma metodología” (ANEP, 2016b). De alguna manera, dadas las ambigüedades y vacíos que deja la propia OCDE en el informe PISA 2015 sobre el efecto de las modificaciones implementadas en el 2015, el país también se reconoce afectado en la dificultad para interpretar las tendencias recientes.

Lamentablemente, la ausencia de una evaluación nacional en años de estudio próximos a la población objetivo de PISA no permite establecer una instancia de comparación con estos diagnósticos divergentes. El **recuadro 9** resume las características de las evaluaciones nacionales de la calidad educativa de Uruguay.

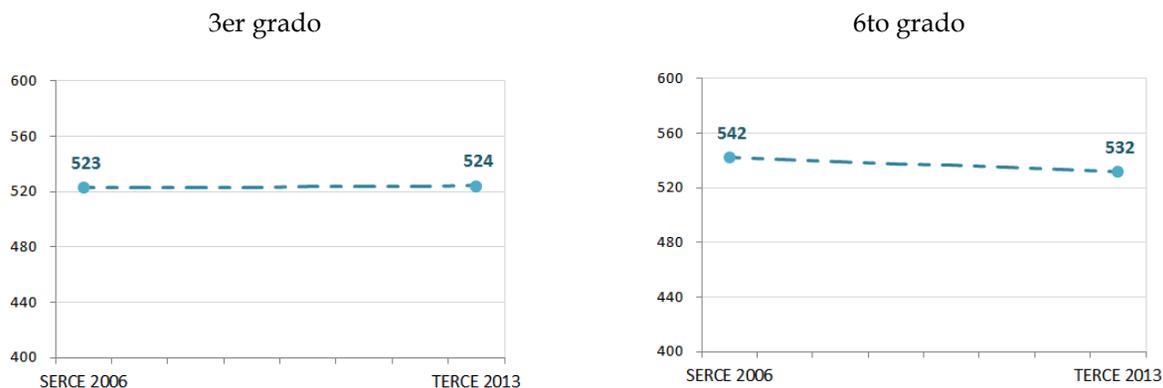
Recuadro 9. Las evaluaciones nacionales de la calidad educativa en Uruguay

Uruguay implementó desde el año 1996 el Programa de Evaluación de Aprendizajes, con una prueba muestral a estudiantes de 6to grado de la primaria. Las evaluaciones eran muestrales y se aplicaron con una frecuencia trienal. En el año 2005 se modificó la metodología de la evaluación, incorporando la Teoría de Respuesta al Ítem como marco metodológico para el diseño de las evaluaciones y el escalamiento de los resultados, lo que implicaba discontinuar la serie respecto de años anteriores. En el diseño de la prueba se tomaron recaudos para contemplar la posibilidad de comparar la aplicación con años anteriores, a partir de una muestra de ítems comunes. En los años siguientes, el país introdujo una innovación en la experimentación con evaluaciones en línea para el nivel primario, en el marco del plan Ceibal de entrega de una computadora por alumno para el nivel primario.

La reciente creación del Instituto Nacional de Evaluación Educativa (INEEd), en funciones desde el año 2013, introduce un nuevo espacio institucional para el desarrollo de las políticas de evaluación educativa en el país, junto con la División de Investigación, Evaluación y Estadística de la Administración Nacional de Educación Pública. Actualmente, el INEEEd se encuentra en proceso de desarrollo de la evaluación “Aristas”, que reemplaza a la evaluación nacional. La implementación de Aristas se propone producir información sobre los desempeños en lectura y matemática de los estudiantes de tercero y sexto de primaria, y de tercero de educación media, con foco en los conocimientos y capacidades. Además, busca producir información sobre habilidades socioemocionales, clima escolar, la convivencia y la participación en los centros.

Las pruebas SERCE-TERCE, implementadas por LLECE-UNESCO, evaluaron a los estudiantes de nivel primario. Estos resultados permiten aproximarse a los desempeños de los estudiantes uruguayos en el contexto de los países latinoamericanos. Si bien sus tendencias no son comparables con las de PISA, su análisis permite identificar señales de la situación de los aprendizajes del nivel primario en el país.

Gráfico 83. Resultados SERCE- TERCE en lectura. Uruguay. 2006-2013

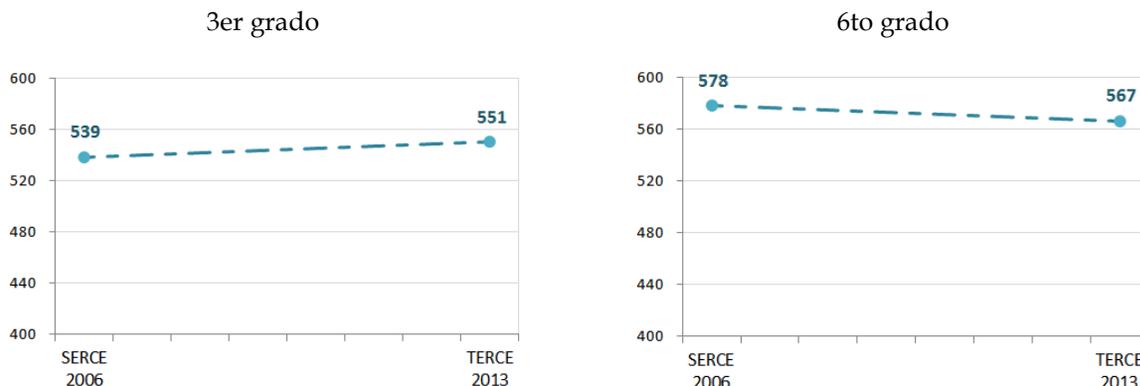


Unidad de medida: Puntaje promedio

Estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados SERCE - TERCE.

Gráfico 84. Resultados SERCE- TERCE en matemática. Uruguay. 2006-2013



Unidad de medida: Puntaje promedio

Estandarizado con media = 500 y desvío estándar = 100

Fuente: CIPPEC, sobre la base de resultados SERCE - TERCE.

En las evaluaciones TERCE, los resultados de aprendizaje de Uruguay posicionaron al país en lugares superiores. Obtuvo en matemática resultados muy por encima del promedio de países evaluados, ocupando el segundo lugar en 6to grado (567 puntos) y el tercero en 3er grado (551 puntos): 520 puntos promedio lectura y matemática en 6to grado (en la región el puntaje promedio fue de 509). En lectura, los resultados también fueron altos, ocupando el tercer puesto después de Chile y Costa Rica en ambos grados. En ciencias, área evaluada sólo en 6to grado, alcanzó los segundos resultados más altos, luego de Colombia.

En la comparación en el tiempo, Uruguay se destaca por haber sido uno de los países que menos creció en sus desempeños entre 2006 y 2013. En 6to grado, las diferencias son negativas: los resultados de TERCE fueron más bajos que el SERCE en las tres áreas evaluadas, situación que

comparte con Costa Rica y en menor medida con Paraguay. En 3er grado, los resultados muestran una leve mejora, muy por debajo del crecimiento promedio de los países: en lectura una mejora de 1,5 puntos en un contexto de incremento de 18,5 puntos en el promedio de países. En matemática, una mejora de 12 puntos, por debajo del promedio de países que incrementó sus puntajes en 31 puntos.

Resulta complejo obtener conclusiones que sintetizen las tendencias perfiladas. El diagnóstico más recurrente es el de un escenario de estabilidad en los desempeños de los estudiantes en los últimos 10 años. Ciertas tendencias contrapuestas arrojan ciertas dudas sobre cuáles son las áreas que se han visto más fortalecidas en el período. El crecimiento intenso que muestra PISA en sus resultados originales en las tres áreas, con particular incidencia en lectura, no parece verse reflejado en otras evaluaciones, lo que induciría a pensar que su diagnóstico no es compartido con otras evaluaciones. El propio reescalamiento de resultados que realiza PISA con la metodología implementada en 2015 constituye una alerta al respecto.

Resumen de datos educativos y sociales destacados de Uruguay

Cobertura

1. El país tiene altas tasas históricas de escolarización, en comparación con el resto de la región: para el año 2000, la tasa neta de escolarización secundaria se situaba en el 74% (SITEAL), solamente por debajo de Argentina (SITEAL) y Chile (SEDLAC) pero marcadamente por encima del promedio de los 7 de 66% (Circa 2014). La tasa de analfabetismo de la población de 15 años y más era, ya para 1980, del 5,4%, la más baja entre los países estudiados.
2. Sin embargo, en la última década la expansión de la educación secundaria muestra una importante desaceleración: entre el 2001 y el 2014, la tasa neta de escolarización del nivel se incrementó sólo 4,4 puntos. Este leve crecimiento se concentró principalmente en los sectores más desfavorecidos: para el quintil I el incremento de la tasa fue del 55% al 73%.
3. Logró avanzar en la expansión del nivel inicial a lo largo del período estudiado: entre 2000 y 2014 la tasa de crecimiento de la matrícula inscripta en el nivel fue del 33,5%, incorporando a 23 mil estudiantes nuevos. Este crecimiento es similar al total de América Latina y el Caribe (24,0%) y superior a Argentina, Brasil y Chile.

Trayectorias de los alumnos

1. Sus tasas de repitencia al año 2014 son unas de las más altas entre los países estudiados en el nivel primario (4,6%), superado solamente por Brasil. Durante el período estudiado, la repetición disminuyó en primaria (al año 2002 era de 8,5%).
2. La población de 15 años presenta un 35% de estudiantes que repitieron de año al menos una vez. A su vez, la brecha en los resultados de las evaluaciones entre los estudiantes que repitieron al menos una vez y los que nunca lo hicieron es la más alta de los 7 países, alcanzando 99 puntos en ambas áreas (lectura y matemática).

Resultados de Evaluaciones

1. En los resultados de equidad, las brechas en los resultados en PISA entre los cuartiles de mayor y menor nivel socioeconómico se situaron en torno a los 100 puntos entre los años 2000 y 2012. Sin embargo, entre 2012 y 2015 se produjo una disminución de la misma de 8 puntos. Fueron las segundas más altas de la región, detrás de Perú. En promedio, en el 2015 la brecha por cuartiles en los países evaluados era de 84 puntos.

2. Considerando la distribución del indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) entre escuelas, un estudio desarrollado sobre la base de datos del año 2015 (Vázquez, 2016) muestra que Uruguay es el país con menores niveles de segregación educativa en función del nivel socioeconómico dentro del universo de países estudiados.

Contexto

1. El indicador de pobreza construido por PISA (*índice del status económico, social y cultural*, ESCS por sus siglas en inglés) perfila un leve mejoramiento en las condiciones de vida de la población de 15 años entre 2003 y 2015. Otros indicadores de pobreza muestran una marcada mejoría: por ejemplo, según fuentes oficiales el porcentaje de población bajo la línea de pobreza descendió del 31,0% al 9,7% en esos años, y el de indigencia se redujo del 1,1% al 0,3%. Otros indicadores, como el porcentaje de población con Necesidades Básicas Insatisfechas calculado por SEDLAC, también muestran una mejora significativa. Estas diferencias entre los indicadores de PISA y los basados en encuestas de hogar introducen el interrogante acerca de la existencia de cierto sesgo en la muestra de escuelas evaluadas.
2. Esta evolución de la pobreza es consistente con el comportamiento de los indicadores macroeconómicos: el inicio de la década del 2000 está marcado por una profunda crisis. El año 2002 es el de mayor contracción de la economía, con una caída anual del PBI del 7%, la más baja de los últimos 25 años. En cambio, la segunda mitad de la década está marcada por un crecimiento sostenido del PBI: entre el 2006 y el 2014 el PBI per cápita se incrementó un 69% a valores de paridad de compra (IMF-WEO).
3. En estos años se mantuvo relativamente estable la distribución social de la riqueza, medida a través del coeficiente de Gini, con un leve descenso del 0,46 en el año 2002 al 0,41 en el año 2014. Sin embargo, Uruguay es, entre los países estudiados, el que manifiesta menores niveles de desigualdad en la distribución del ingreso medido a través del coeficiente de Gini: el valor promedio de este coeficiente en los países estudiados es de 0,47 (SEDLAC).
4. Uruguay también manifiesta altos niveles de empleo formal (el 12,5% de los trabajadores no poseen cobertura de pensión, contra el 33% promedio de los siete países), niveles medios de desempleo (7,6% en el país y 6,7% en el promedio), y muy altos niveles de cobertura de salud (el 98% de la población ocupada urbana posee cobertura de salud, la más alta de los 7 países, cuyo promedio es de 70,3%)

Inversión educativa

1. El país aumentó fuertemente su inversión en el sistema educativo en la última década. Medido como porcentaje del PBI, el gasto en educación creció del 2,1% en el 2003 al 4,3% en 2011. Este crecimiento es, en puntos porcentuales, el más alto de América Latina en el periodo estudiado. Gracias a este crecimiento, al año 2011 alcanza niveles de inversión (medidos como % del PBI) en un rango intermedio en la región (el promedio de 16 países de la región es 5,0 en valores *circa* 2013).
2. Asimismo, el gasto en educación como porcentaje del gasto público total, también tuvo un incremento. En 2000 alcanzaba el 7,6% y para el año 2013 era de 12,3%.

Conclusiones

El largo y profundo trabajo de sistematización de las evaluaciones de la calidad educativa en América Latina, con especial énfasis en la prueba PISA, nos deja llenos de interrogantes. ¿Qué países mejoraron la calidad de los aprendizajes medidos por las pruebas? ¿Qué consistencia tiene su evolución dispar en pruebas paralelas en el tiempo? ¿Qué tan fiables son las metodologías de las pruebas para captar lo que ocurre en los sistemas educativos? ¿Cómo se han adaptado los países a las pruebas y hasta qué punto no se ha convertido esa adaptación en un riesgo para la estricta comparabilidad intertemporal de los resultados? Estas preguntas encuentran numerosas respuestas parciales en este documento.

Sabemos que algunos países han tenido ciclos de mejora más claros que otros. Perú muestra mejoras en todas las pruebas analizadas, con gran magnitud desde un punto de partida muy bajo hasta llegar más cerca del promedio regional. Además consiguió en la etapa estudiada (2000 a 2015) importantes avances en la inclusión y la trayectoria de los alumnos. Con grandes deudas en el terreno de las desigualdades, Perú muestra un camino de mejoramiento educativo notable en el siglo XXI.

Partiendo de una situación muy distinta, Chile parece haber mejorado en los años 2000, aunque luego encuentra una etapa de estancamiento que lo mantiene por encima de los demás países analizados. En el nivel primario las mejoras en la prueba de la UNESCO se destacan en matemática, con resultados claramente por encima de la región. La evolución en las pruebas nacionales, sin embargo, no es tan clara y presenta vaivenes. En el nivel secundario las mejoras en PISA se dieron sobre todo en el ciclo 2000 a 2006, mientras los resultados en las pruebas nacionales son en cierto sentido contradictorios, con mejoras en matemática y un retroceso reciente en lectura.

Estos logros se dieron en un período de crecimiento del acceso y mejora de las trayectorias educativas de los alumnos, que se encuentran en la actualidad claramente por encima del resto de los países de la región, por ejemplo en las tasas de terminalidad de la educación secundaria.

Brasil parece tener una marcada diferencia en sus avances en la calidad educativa en el nivel primario y medio. En el nivel primario el país mejora en las pruebas internacionales de la UNESCO, especialmente en matemática. Las pruebas nacionales también indican un claro ciclo de mejora desde 2005 hasta 2015. En el nivel medio, la prueba PISA no muestra grandes avances, salvo en la primera etapa, entre 2000 y 2006, con una mejora leve que quizás se explica por una mayor importancia asignada a la prueba y una consecuente reducción de la no respuesta. Cabe destacar que estos resultados se enmarcan en un ciclo de gran expansión del acceso a la educación secundaria y de reducción de la repetición en todo el sistema.

México también tiende a una estabilidad en la educación secundaria con mejoras en el nivel primario. Lo más notable es la mejora en matemática en el nivel primario, tanto en las pruebas internacionales como, especialmente, en las nacionales. En PISA México oscila con vaivenes y una mejora leve en matemática. En las pruebas nacionales se observan mejoras especialmente en algunos años particulares. También México logra en este período un gran incremento de las tasas de acceso al nivel medio, desde puntos de partida muy bajos.

En Colombia, los resultados muestran estabilidad tanto en las pruebas internacionales de PISA (luego de ser reescaladas para ser comparables en el tiempo) como en las nacionales, donde hay vaivenes difíciles de explicar. En el nivel primario las mejoras son leves en las pruebas internacionales y no se observan con claridad en las pruebas nacionales. En cambio, se constata un

gran avance del país en el aumento de las tasas de escolaridad y reducción de la sobreedad en el nivel secundario.

En Uruguay, se observa un ciclo de estabilidad en los resultados. Con el reescalamiento de los resultados de PISA no se observan cambios en el período 2006-15 y no existen pruebas nacionales para contrastar estos avances. En las pruebas nacionales e internacionales de nivel primario el país también muestra estancamiento, mientras el resto de la región presentó avances.

En Argentina hay mejoras en el nivel primario tanto en las pruebas nacionales como internacionales (más en matemática que en lengua). También en las pruebas nacionales se constatan mejoras en lengua, mientras en las pruebas PISA hay más estabilidad hasta 2012 y muchas preguntas abiertas por la imposibilidad de comparabilidad de la prueba 2015.

Como se puede observar, las trayectorias no son lineales y muchas veces muestran tendencias contrapuestas según áreas de enseñanza evaluadas o niveles educativos. Esto requiere estudios más profundos sobre las diferencias dentro de cada sistema educativo, tomando en cuenta la gran diversidad territorial y la descentralización de la gestión escolar que tienen casi todos los países analizados. La investigación sobre los niveles subnacionales de gobierno y sobre las prácticas de enseñanza son dos caminos posibles para continuar esta investigación.

El estudio también demuestra que es necesario tener una mirada más amplia de la evaluación de la calidad educativa. No basta con leer las tablas de PISA. Muestran un recorte muy parcial y ni siquiera han sido claras para exponer la validez de la evolución de los resultados. La edición 2015 de la prueba tuvo cambios metodológicos considerables que afectaron la comparabilidad de los resultados, especialmente en América Latina. Esto debió ser mejor aclarado por los documentos de PISA. Países como Uruguay y Colombia fueron exhibidos como casos de mejora entre 2012 y 2015 cuando la propia prueba PISA al reescalar los resultados y hacerlos estrictamente comparables muestra que no hubo mejora.

Hay que mirar en detalle qué validez tiene PISA para entender los sistemas educativos. Eso es lo que intentó mostrar el documento, destacando el gran valor comparado que permite no solo la prueba sino la riqueza de sus cuestionarios complementarios. Pero también es importante leer esos resultados a la luz de otras evaluaciones, intentando encontrar tendencias comunes y preguntando sobre las disidencias.

Todavía es necesario estudiar más la validez que aportan las evaluaciones estandarizadas de los aprendizajes. Hay que ser cuidadosos y no confiar excesivamente en sus resultados como reflejo de sistemas educativos complejos y multidimensionales.

Es clave multiplicar, no reducir, las evaluaciones y mediciones de los sistemas educativos. Contar con las evaluaciones de la OCDE, UNESCO, otros organismos y los propios países, que han consolidado en los años recientes sistemas de evaluación propios con mucho esfuerzo y, en varios casos, con gran crecimiento institucional acumulado. Estos esfuerzos permiten comprender mejor el funcionamiento de los sistemas educativos para saber cuáles mejoran y, hasta cierto punto, tomar lecciones de las posibles hipótesis que explican esas mejoras.

En definitiva, estos estudios permiten analizar el cumplimiento del derecho a la educación que se materializa en los aprendizajes de los alumnos. Sin datos rigurosos apenas tenemos visiones muy lejanas sobre el cumplimiento de los derechos.

Pero también las evaluaciones han traído un efecto de banalización del debate educativo. Esto no es responsabilidad de quienes diseñan las evaluaciones y los informes, sino de los que utilizan estos datos de manera inconexa o sesgada.

Saber qué países mejoran es un dato central para entender la dinámica de los sistemas educativos a gran escala. Esto requiere mucha responsabilidad. Más aún si se usan esos resultados para tomar lecciones de política educativa. El salto epistemológico entre resultados en las pruebas y posibles políticas que los expliquen es grande aunque no inconmensurable, como hemos mostrado en estudios anteriores (Rivas, 2015). Ese salto es un trabajo serio que no admite tentaciones de favoritismos ideológicos ni atajos estadísticos.

La tarea de la investigación para la toma de decisiones no termina en esta mirada rigurosa de los datos. Entender los sistemas educativos requiere estudiar su historia, sus contextos, sus culturas. La verdadera mirada comparada apenas comienza con las evaluaciones de la calidad. Las explicaciones exigen otro caudal metodológico que sólo unos pocos trabajos han llegado a explorar (Alexander, 2012; Carnoy, 2007).

Para las gestiones educativas de los países, este estudio tiene un valor vital en la comprensión de su propia evolución en las evaluaciones de la calidad y su rol como autoridades de ejecución de las pruebas. Se han logrado grandes avances en la calidad de la implementación de las pruebas, no sin esfuerzos ni linealmente. Todavía existen grandes desafíos para evitar la disparidad de dosis de preparación y presión en la ejecución de las pruebas. Debe lograrse una estabilidad en el tiempo que garantice aplicaciones eficaces y resultados comparables. De lo contrario los resultados pueden tener una dosis de incertidumbre que lleve a confusiones.

Este desafío abarca a las pruebas internacionales y nacionales. Cuanto mayor es la visibilidad de los instrumentos mayor debe ser su calidad institucional y su protección de cualquier uso político de corto plazo.

Aprender sobre la dinámica de los sistemas educativos es una pieza vital del proceso de mejora. Requiere datos rigurosos, comparables en el espacio y el tiempo; investigaciones más profundas que miren las culturas, las prácticas y la historia; miradas sensatas que busquen la verdad dentro de la racionalidad científica, evitando los peligrosos sesgos ideológicos que recorren los debates educativos y políticos de cada país. Por esa vía será posible tener más conocimientos para garantizar el derecho a la educación.

Bibliografía

Alexander, R. (2012). Moral Panic, Miracle Cures and Educational Policy: what can we really learn from international comparison? *Scottish Educational Review*, 44 (2012).

Agencia de Calidad de la Educación (2014). Informe Nacional Resultados Chile PISA 2012. Santiago de Chile.

Agencia de Calidad de la Educación (2015). *PISA 2015. Programa para la Evaluación Internacional de Estudiantes OCDE*. Santiago de Chile.

ANEP (2016a). *Primer informe de resultados*. Montevideo. Administración Nacional de Educación Pública. Consejo Directivo Central. División de Investigación, Evaluación y Estadística. Programa Nacional ANEP-PISA.

ANEP (2016b) *Nota metodológica sobre las estimaciones de los resultados PISA 2015*. Montevideo. Administración Nacional de Educación Pública. Consejo Directivo Central. División de Investigación, Evaluación y Estadística. Programa Nacional ANEP-PISA.

Carnoy, M. (2007). *Cuba's Academic Advantage*. Stanford: Stanford University Press.

Carnoy, M. (2015). *International Test Score Comparisons and Educational Policy: A Review of the Critiques*. Boulder, CO: National Education Policy Center.

Carnoy, M., y Rothstein R. (2013). *What Do International Tests Really Show About American Student Performance?* Economic Policy Institute.

Eivers, E. (2010). PISA: issues in implementation and interpretation. *The Irish Journal of Education*, 38, pp. 94-118.

Fernandez Cano, A. (2016). *A Methodological Critique of the PISA Evaluations*. *Relieve*, 22(1). Disponible en: <http://dx.doi.org/10.7203/relieve.22.1.8806>.

Freitas, P.; Catela N.; Balcão Reis, A.; Seabra, C. y Ferro, A. (2015). *Correcting for sample problems in PISA and the improvement in Portuguese students' performance*. *Assessment in Education: Principles, Policy & Practice*, 23 (4), pp. 456-472.

Goldstein, H. (2004). *International comparisons of student attainment: some issues arising from the PISA study*. *Assessment in Education*, 11 (3), pp. 319-30.

INEE (2016). *México en PISA 2015*. México: INEE.

Jerrim, J. (2013). *The Reliability of Trends over Time in International Education Test Scores: Is the Performance of England's Secondary School Pupils Really in Relative Decline?* *Journal of Social Policy*, 42, pp. 259-279.

Jerrim, J. y Micklewright, J. (2012). *Socioeconomic gradients in children's cognitive skills: Are cross-country comparisons robust to who reports family background?* Department of Quantitative Social Science, Institute of Education, University of London.

Klein, R. (2011). Un reanálisis de los resultados del PISA: Problemas de comparabilidad. *Ensaio: Avaliação e Políticas Públicas em Educação*, 19 (73).

Kreiner, S. (2011). *Is the Foundation under PISA Solid? A Critical Look at the Scaling Model Underlying International Comparisons of Student Attainment*. Department of Biostatistics Research Report 11/1.

Lacasa, J. M. (2011). *PISA 2009: Una segunda mirada*. Madrid: Confederación Española de Centros de Enseñanza.

Meyer, H.-D. y Benavot, A. (2013). *PISA, Power, and Policy: the Emergence of Global Educational Governance*. Oxford: Symposium Books.

Ministerio de Educación (2016). *Brasil no PISA 2015: análises e reflexões sobre o desempenho dos estudantes brasileiros*. São Paulo: Fundación Santillana.

Ministerio de Educación (2017). *El Perú en PISA 2015. Informe nacional de resultados*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.

O'Leary, M. (2001). *The effects of age-based and grade-based sampling on the relative standing of countries in international comparative studies of student achievement*. *British Educational Research Journal*, 27, pp. 187-200.

OCDE (2003). *PISA 2003 Technical Report*. París: OCDE.

OCDE (2006). *PISA 2006 Technical Report*. París: OCDE.

OCDE (2009). *PISA 2009 Technical Report*. París: OCDE. Disponible en: <http://dx.doi.org/10.1787/9789264167872-en>

OCDE (2016a). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. París: OCDE. Disponible en: <http://dx.doi.org/10.1787/9789264266490-en>.

OCDE (2016b). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. París: OCDE. Disponible en: <http://dx.doi.org/10.1787/9789264267510-en>.

OCDE (2016c). *Country note. Results from PISA 2015-Brazil*. París: OCDE.

OCDE (2016d). *Country note. Results from PISA 2015-Colombia*. París: OCDE.

OCDE (2016e). *Country note. Resultados de PISA 2015-México*. París: OCDE.

OCDE (2016f). *PISA 2015 Technical Report*. París: OCDE. Disponible en: <http://www.OCDE.org/pisa/data/2015-technical-report/>

OCDE (2016g). *Country note. Resultados de PISA 2015-Uruguay*. París: OCDE.

OCDE (2016h). *Competencias en Iberoamérica: análisis de PISA 2012*. Lima: Santillana.

OCDE (2017). *Main survey school sampling preparation manual. Overview*. París: OCDE. Disponible en: <https://www.oecd.org/pisa/pisaproducts/MAIN-SURVEY-SCHOOL-SAMPLING-PREPARATION-MANUAL.pdf>

Pereyra, M. A., Kotthoff, H.-G. y Cowen, R. (2011). *PISA Under Examination: Changing Knowledge, Changing Tests, and Changing Schools*. Holanda: Sense Publishers, Cese.

Prais, S. J. (2003). *Cautions on OCDE's Recent Educational Survey (PISA)*. *Oxford Review of Education*, 29 (2), pp. 139-163.

Rivas, A. (2015). *América Latina después de PISA. Lecciones aprendidas de la educación en siete países (2000-2015)*. Buenos Aires: CIPPEC, Natura, Instituto Natura.

Scasso, M. y Massé, G. (2016) "Desarrollo de alternativas para la estimación de demanda poblacional sobre el sistema educativo por edad simple". Documento elaborado por UNICEF LACRO y la Asociación Civil Educación Para Todos. Disponible en: <http://allinschool.org/wp-content/uploads/2016/07/LACRO-Final-version-Desarrollo-de-alternativas-para-la-estimacion-de-demanda-poblacional-final.pdf>

Secretaría de Evaluación Educativa (2016) Operativo Nacional de Evaluación (ONE) 2013. Diagnóstico y consideraciones metodológicas necesarias para el análisis y difusión de sus resultados. Ministerio de Educación y Deportes. República Argentina

Seth, M. (2007). *Fiebre Educativa. Sociedad, política, y el anhelo de conocimiento en Corea del Sur*. Buenos Aires: Prometeo.

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*. Vol. 75 No. 3, p. 417-453.

Solano-Flores, G. y Milbourn, T. (2016). *Capacidad evaluativa, validez cultural y validez consecuencial en PISA*. *Relieve*, 22(1).

Wuttke, J. (2007). Uncertainties and Bias in PISA. PISA according to PISA. Does PISA keep what it promises? Viena: Hopmann, Brinek, Retzl (eds.), pp. 241-263. Disponible en: <https://ssrn.com/abstract=1159042>.

Anexo I Datos de contexto, indicadores sociales y educativos de los países

Cuadro 7. Períodos presidenciales. Países seleccionados. 2000-2015

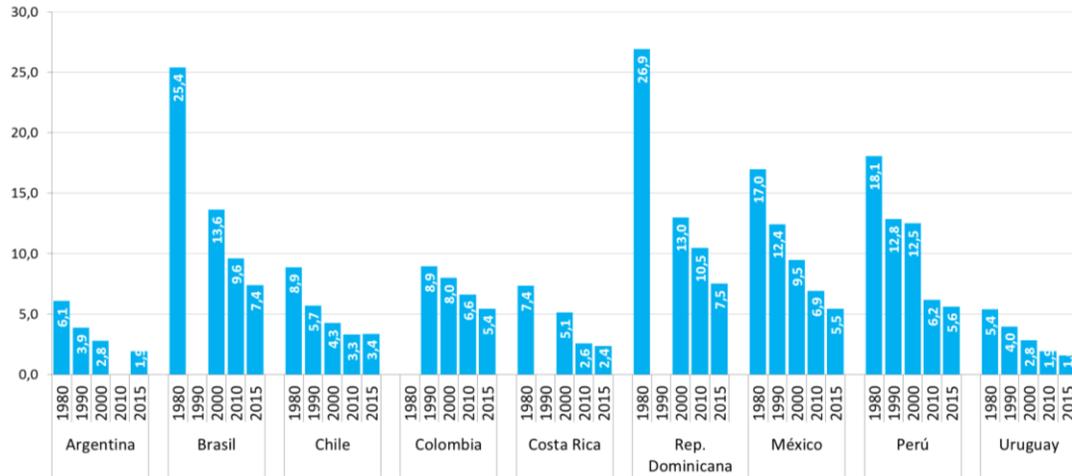
Pais	Argentina	Brasil	Chile	Colombia	México	Perú	Uruguay
1995-99			E. Frei Ruiz-Tagle (marzo 1994)				
2000	F. De la Rúa (diciembre 1999)	F. Henrique Cardoso (enero 1995)		A. Pastrana (junio 1998)		V. Paniagua (nov. 2000)	
2001							
2002	E. Duhalde (enero 2002)		R. Lagos (marzo 2000)		V. Fox (diciembre 2000)	A. Toledo Manrique (julio 2001)	J. Batlle (marzo 2000)
2003							
2004	N. Kirchner (mayo 2003)						
2005							
2006		I. da Silva (enero 2003)		A. Uribe (agosto 2002)			
2007			M. Bachelet (marzo 2006)				T. Vázquez (marzo 2005)
2008					F. Calderón (diciembre 2006)	A. García Pérez (julio 2006)	
2009							
2010	C. Fernández de Kirchner (diciembre 2007)						
2011			S. Piñera (marzo 2010)				
2012							J. Mujica (marzo 2010)
2013		D. Rousseff (enero 2011)		J. M. Santos (agosto 2010)		O. Humala (julio 2011)	
2014							
2015			M. Bachelet (marzo 2014)		E. P. Nieto (dic. 2012)		
2016	M. Macri (diciembre 2015)	M. Temer (agosto 2016)				Kuczynski Godard (julio 2016)	Tabaré Vázquez (marzo 2015)
2017							

Nota: actualizado al 13 de septiembre de 2017.

Fuente: CIPPEC.

Indicadores de cobertura

Gráfico 85. Tasa de analfabetismo en la población de 15 y más años

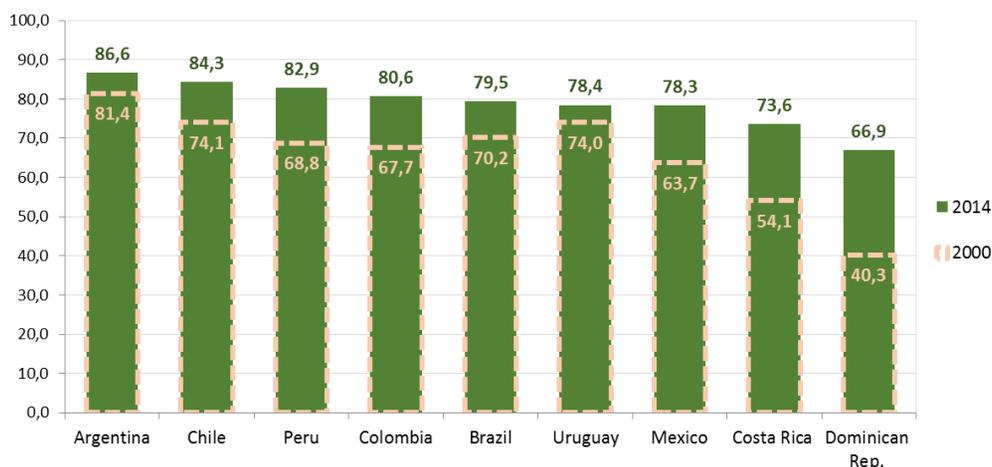


■ Tasa de analfabetismo en la población de 15 y más años

Los datos de 1980 corresponden a 1980 para Argentina, Brasil y México; 1981 para Perú y Rep. Dominicana; 1982 para Chile; 1984 para Costa Rica; y una estimación basada en los valores de 1975 y 1985 para Uruguay. Los datos de 1990 corresponden a 1990 para México; 1991 para Argentina; 1992 para Chile; 1993 para Colombia y Perú, y una estimación basada en valores de 1985 y 1996 para Uruguay. Los datos de 2000 corresponden a 2000 para Brasil, Costa Rica y México; 2001 para Argentina; 2002 para Chile y Rep. Dominicana; una estimación basada en los valores de 1996 y 2004 para Colombia; de 1993 y 2004 para Perú, y de 1996 y 2006 para Uruguay. Los datos de 2010 corresponden a 2010 para Brasil, Colombia, México, Rep. Dominicana y Uruguay, 2011 para Chile y Costa Rica, y 2012 para Perú. Los datos de 2015 corresponden a 2015 para todos los países.

Fuente: CIPPEC, sobre la base de UIS-UNESCO.

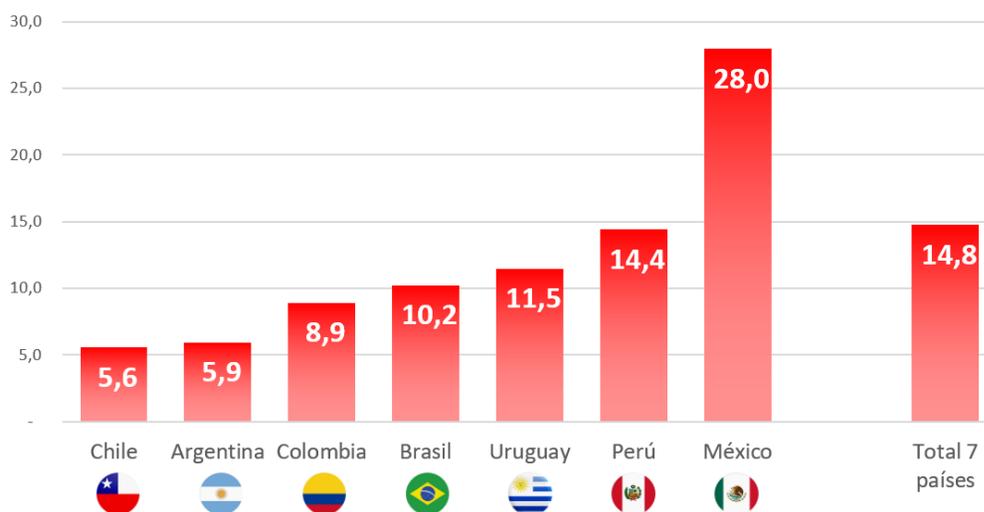
Gráfico 86. Tasa neta de escolarización en el nivel secundario. Países seleccionados. 2000 y 2012



Nota: los datos 2000 corresponden a 2000, excepto para Brasil, Colombia y Uruguay que corresponden a 2001. Los datos 2014 corresponden a 2014, excepto para Chile que corresponden a 2013.

Fuente: CIPPEC, sobre la base de SITEAL, IPE-UNESCO y OEI (Argentina, Brasil México y Uruguay), SEDLAC-Banco Mundial (Chile, Colombia, Rep. Dominicana), Dirección de Planificación Institucional, Ministerio de Educación Pública (Costa Rica) y Estadística de la Calidad Educativa - ESCALE (Perú).

Gráfico 87. Porcentaje de población en edad de asistir a la educación secundaria (*) que se encuentra fuera de la escuela. Países seleccionados. Circa 2015()**

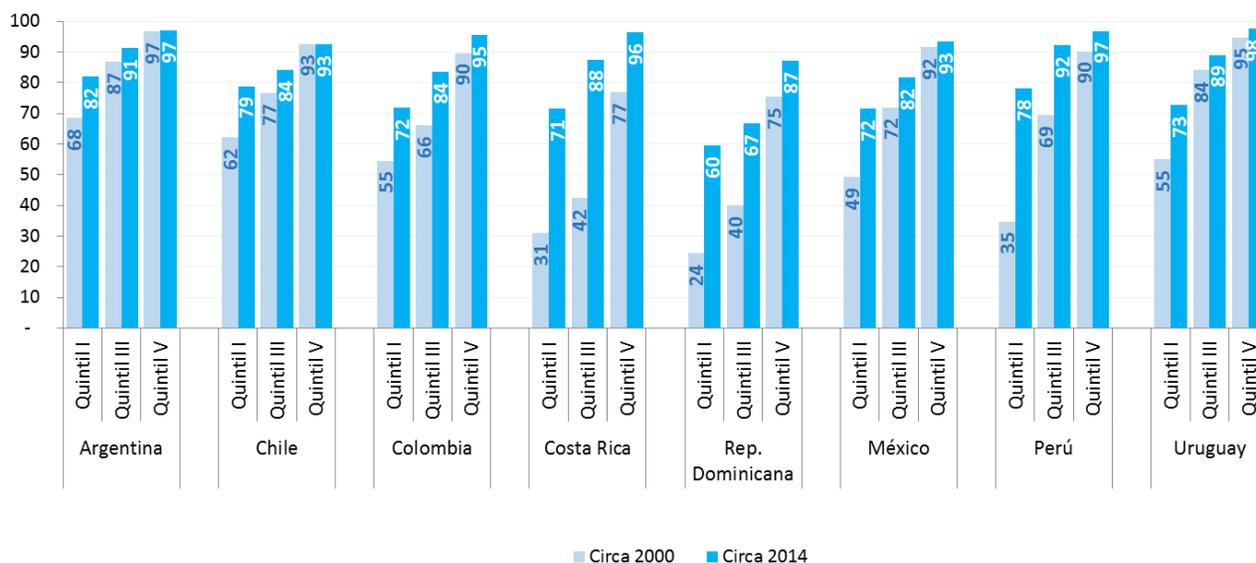


(*) El cálculo considera a la población en edad teórica de asistir a la educación secundaria que no se encuentra asistiendo al nivel secundario ni primario. Se toma como referencia al rango de edad que corresponde al nivel secundario según cada estructura de niveles.

(**) Los datos corresponden a 2015 para todos los países excepto Argentina y Uruguay (2014); México (2012).

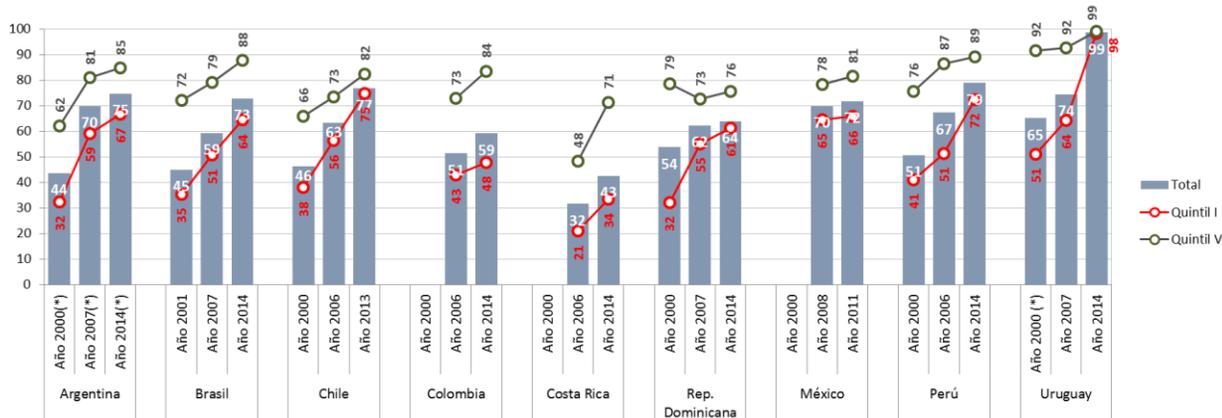
Fuente: CIPPEC, sobre la base de UIS data centre

Gráfico 88. Tasa neta de escolarización en el nivel secundario por quintiles de ingreso. Países seleccionados. Circa 2000 y 2014



Nota: los datos circa 2000 corresponden al año 2000, excepto para Colombia que corresponden a 2001 y Rep. Dominicana que corresponde a 2002. Los datos circa 2014 corresponden al año 2014, excepto para Chile que corresponden al 2013. Los datos de Argentina 2000 y 2014, y de Uruguay 2000, corresponden sólo a zonas urbanas. Los datos de Brasil no se muestran debido a inconsistencias en el valor del indicador. Fuente: CIPPEC, sobre la base de SEDLAC.

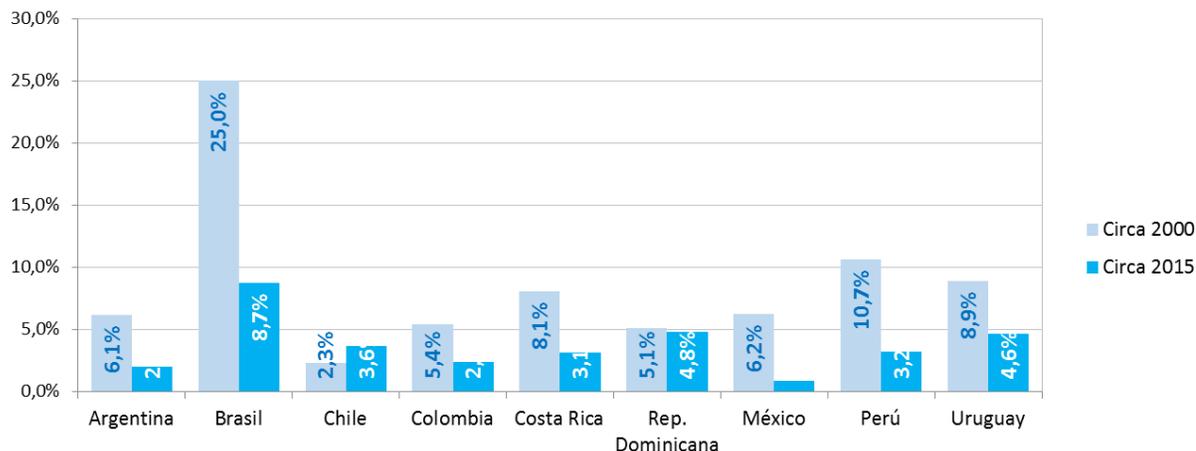
Gráfico 89. Tasa bruta de escolarización (*) de la población de 3 a 5 años de edad. Países seleccionados. Circa 2000, 2007 y 2014



(*) Los datos de Argentina 2000-2014 y Uruguay 2000 corresponden a zonas urbanas. Los datos de Costa Rica previos al año 2005 no son comparables con los de años posteriores, se excluyen del análisis. En México no se dispone de información entre los años 1998 y 2008. En Colombia se consideran los resultados de la Gran Encuesta Integrada de Hogares (GEIH), implementada desde el año 2008. Fuente: CIPPEC, sobre la base de SEDLAC.

Indicadores de trayectoria

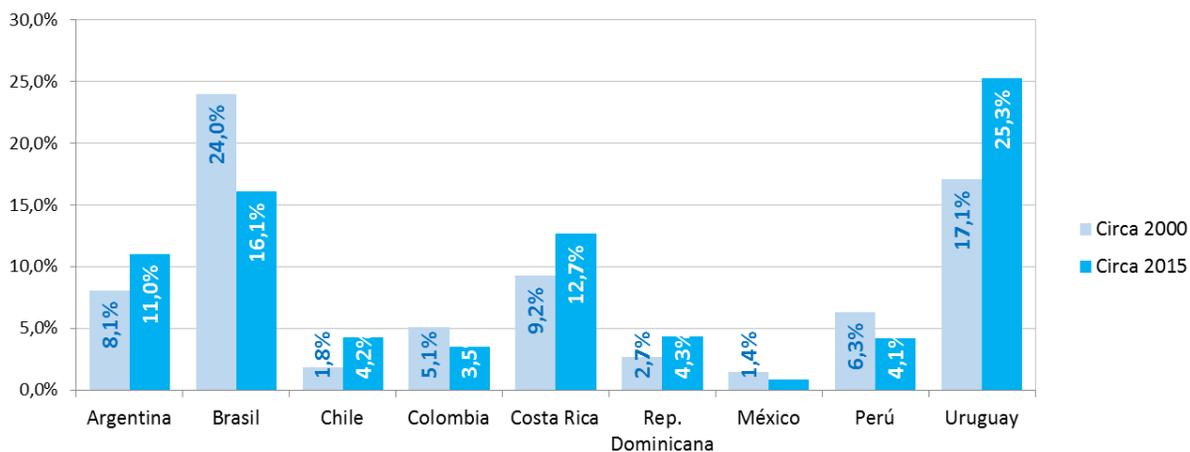
Gráfico 90. Evolución del porcentaje de alumnos repitentes, nivel primario (CINE 1). Países seleccionados. 2001 y 2015



Nota: los datos circa 2000 corresponden al año 2000, excepto para Colombia que corresponden a 2001 y Chile que corresponde a 2002. Los datos circa 2015 corresponden al año 2015, excepto para Brasil que corresponden al 2011, Perú que corresponden a 2014 y Uruguay que corresponden a 2013.

Fuente: CIPPEC, sobre la base de UIS-UNESCO.

Gráfico 91. Evolución del porcentaje de alumnos repitentes, nivel secundario básico (CINE 2). Países seleccionados. 2001 y 2015

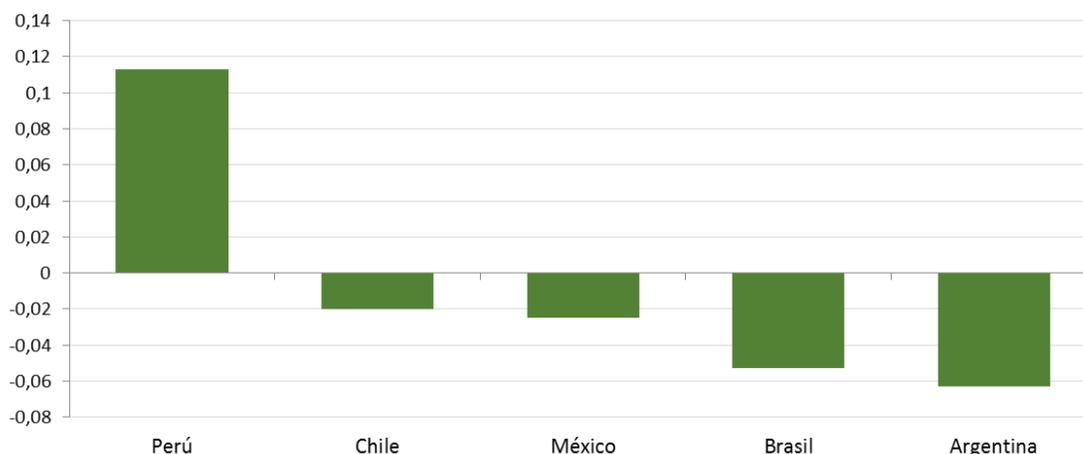


Nota: los datos circa 2000 corresponden al año 2000, excepto para Chile que corresponde a 2002. Los datos circa 2015 corresponden al año 2015, excepto para Brasil que corresponden al 2011, Perú que corresponden a 2014 y Uruguay que corresponden a 2013.

Fuente: CIPPEC, sobre la base de UIS-UNESCO.

Indicadores de equidad

Gráfico 92. Variación en los niveles de segregación escolar por nivel socioeconómico. Índice de Brecha por centiles (CGI)



Fuente: Vázquez (2016), sobre la base de OCDE, 2016.

Indicadores de contexto

Cuadro 8. Resumen de indicadores económicos y sociales. Países seleccionados circa 2015 y variación circa 2000 - circa 2015

	PBI per cápita ajustado por PPP (dólares internacionales corrientes)		Pobreza Extrema * (% de personas debajo de la línea de pobreza extrema)		Pobreza Moderada * (% de personas debajo de la línea de pobreza moderada)		Índice de Gini **		
	2015	Variación % 2000-2015	circa 2015	Variación circa 2000 - circa 2015	circa 2015	Variación circa 2000 - circa 2015	2014	Variación 2000-2014	
Argentina	22.553,58	90,7%	6,3	-1,4	32,2	25,7	41,8	-7,8	Gran mejora
Brasil	15.614,53	71,4%	4,2	-11,1	13,3	-5,5	51,2	-6,6	Mejora moderada
Chile	23.459,56	106,2%	2,5	-3,1	11,7	-2,7	50,4	-4,7	Poca variación
Colombia	13.846,51	109,1%	7,9	-9,8	27,8	-4,9	53,3	-3,1	Poca variación
México	17.534,44	48,9%	20,5	-3,6	29,1	2,0	48,9	-4,5	Mejora moderada
Perú	12.194,70	130,0%	4,3	-10,7	21,8	-4,0	44,1	-5,7	Mejora moderada
Uruguay	21.506,52	113,0%	0,3	-1,2	9,7	-8,6	41,6	-3,1	Poca variación
Promedio	18.101,41	95,6%	6,6	-5,8	20,8	0,3	47,3	-5,1	

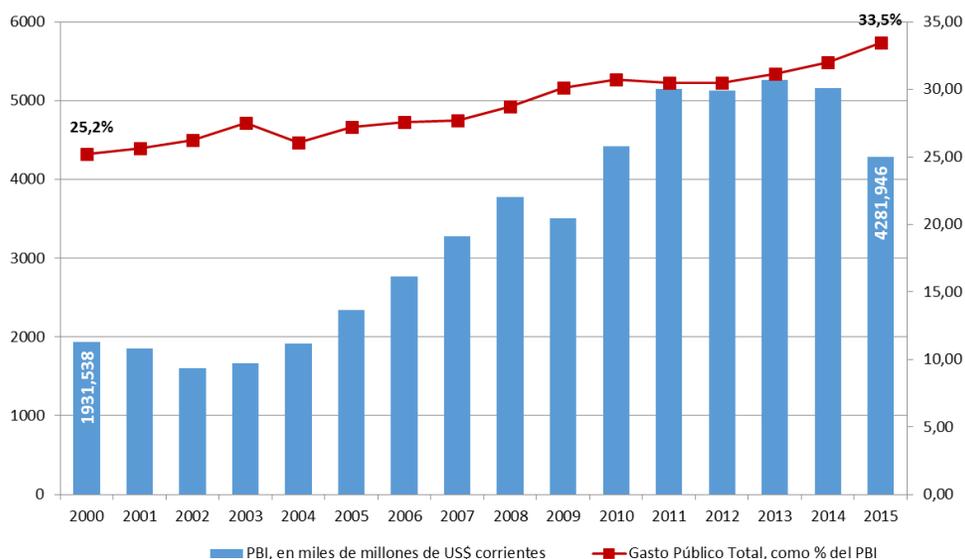
Notas: * Los datos "circa 2000" corresponden al 2000 para Chile, México y Perú; 2001 para Argentina y Brasil; 2002 para Colombia. Los datos "circa 2015" corresponden a 2013 para Chile; 2014 para Brasil, México y Uruguay; 2015 para Argentina, Colombia y Perú.

Los datos para Argentina corresponden a las zonas urbanas.

** Para el caso de Chile, los datos corresponden a 2013. Para el caso de Brasil y Colombia, corresponde a 2001, no 2000.

Fuente: CIPPEC, sobre la base de: PBI per cápita ajustado por PPP: WEO-IMF. Pobreza Extrema y Moderada: SEDLAC, tasas de pobreza oficiales 2000; excepto México (CONEVAL). La pobreza de 2015 Pobreza es de fuentes oficiales: Argentina: INDEC; Brasil: IPEADATA (2014); Chile: CASEN; Colombia: DANE; México: CONEVAL (2014); Perú: INEI; Uruguay: INE. Índice de Gini: SEDLAC-Banco Mundial.

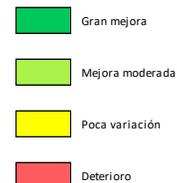
Gráfico 93. PBI y gasto público total como porcentaje del PBI. Países seleccionados. 2000-2015



Fuente: CIPPEC, sobre la base de WEO-IMF y UIS

Cuadro 9. Indicadores sociales complementarios. Países seleccionados, circa 2015 y variación circa 2000-circa 2015

	Esperanza de vida al nacer (años)		Desempleo (como % de la fuerza laboral)		Informalidad Laboral * (% de trabajadores asalariados sin acceso a una pensión tras su retiro)		Cobertura Salud ** (% población ocupada urbana con protección de salud)		Cobertura Pensiones *** (% población ocupada urbana con protección de pensiones)	
	2015	Variación 2000-2015	2015	Variación 2000-2013	2014	Variación 2000 - 2014	Circa 2014	Variación 2000-2014 ***	2014	Variación 2000-2014
Argentina	75,9	2,3	6,5	-10,7	33,0	-5,5	73,2	5,4	53,1	4,0
Brasil	74,5	4,5	6,8	-0,3	22,8	-13,9	67,4	14,0	71,7	13,8
Chile	79,9	3,2	6,2	-3,5	15,5	-7,4	95,5	7,0	69,9	3,2
Colombia	74,3	3,4	8,9	-4,4	38,0	-20,7	48,1	0,7	42,9	8,6
México	76,3	1,9	4,3	2,1	62,1	7,3	47	-1,9	47,8	2,9
Perú	74,9	4,2	6,0	-1,8	52,3	-24,8	63,2	27,7	42,9	9,1
Uruguay	76,5	1,9	7,6	-5,8	12,5	-10,8	98,3	2,1	77	12,0
Promedio	76,1	3,0	6,6	-3,5	33,7	-10,9	70,39	7,9	57,90	7,7



Notas:

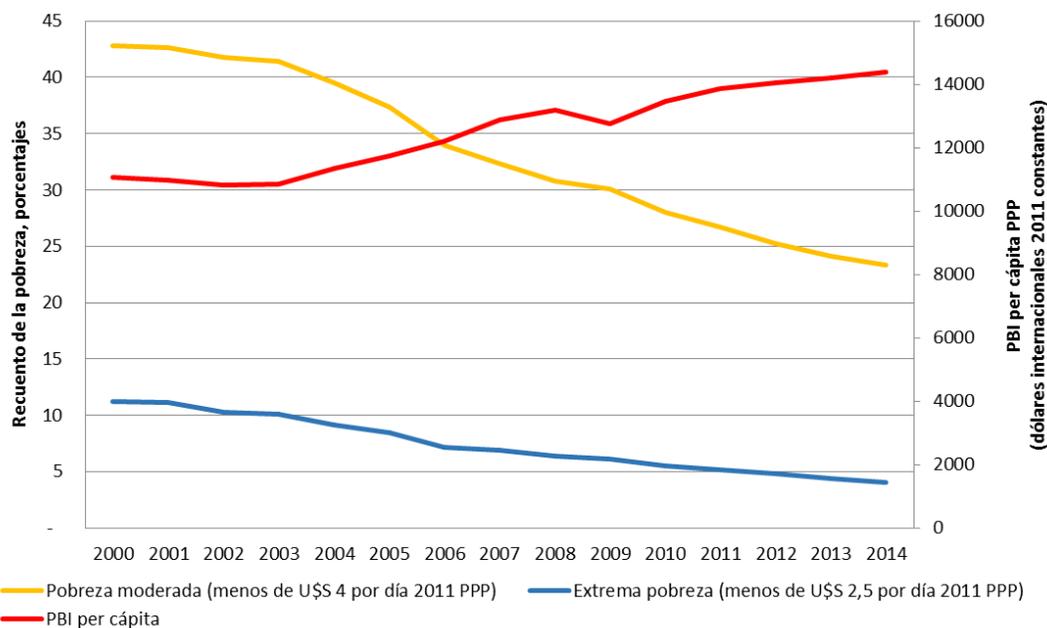
* Los datos 2014 corresponden al año 2014 excepto para México y Chile, donde corresponden a 2013. Los datos 2000 corresponden al año 2000 excepto para Brasil y Colombia donde refieren a 1999, y 2001 para Uruguay

** Los datos 2014 corresponden al año 2014 excepto para Chile, que corresponden al 2011. Los datos “circa 2000” corresponden a 2000 para Argentina, Chile, México, Perú y Uruguay; y 2001 para Brasil y Colombia. No se coloreó la variación uruguaya ya que en 2000, los datos eran solo del ámbito urbano, mientras que en 2012 eran del ámbito rural y urbano.

*** Los datos 2014 corresponden al año 2014 excepto para México, que corresponden al 2013, y Chile que corresponden al 2011. Los datos 2000 corresponden a 2000 para Chile, Colombia y México; 2001 para Brasil y Uruguay; y 2007 para Argentina y Perú (estos datos no fueron coloreados por falta de comparabilidad).

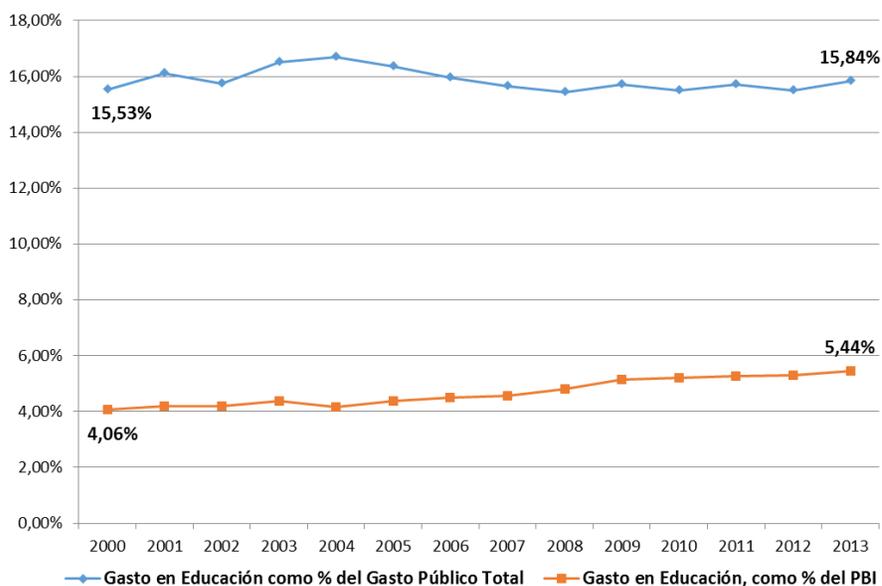
Fuente: CIPPEC, sobre la base de: Esperanza de vida al nacer: Global Health Observatory Data Repository- WHO. Desempleo: WEO-IMF. Informalidad laboral: SEDLAC. Cobertura Salud y Cobertura Pensiones: OIT (2015).

Gráfico 94. Pobreza moderada y extrema (porcentajes) y PBI per cápita (dólares internacionales 2005 constantes). América Latina y el Caribe. 2000-2014



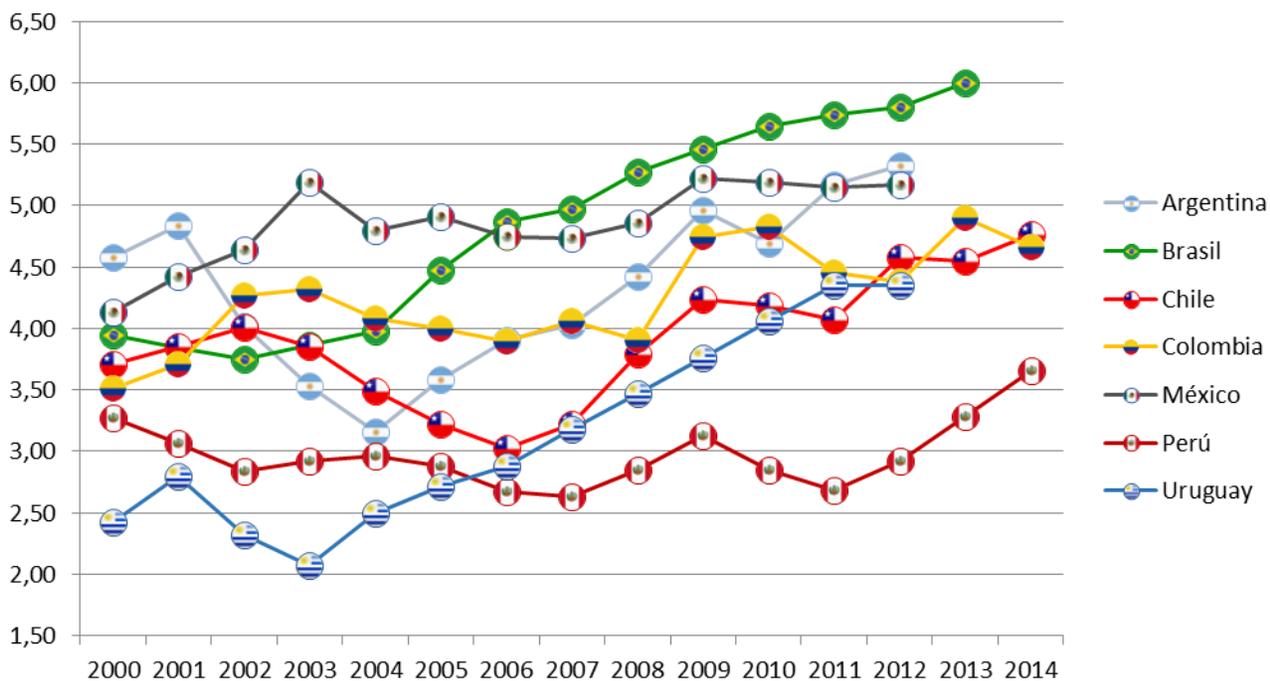
Indicadores de inversión educativa

Gráfico 95. Gasto en educación como porcentaje del PBI y como porcentaje del gasto público total. Promedio de países seleccionados. 2000-2013



Fuente: CIPPEC, sobre la base de: Gasto Público Total: WEO-IMF. Gasto en Educación como porcentaje del PBI: UIS-UNESCO; excepto en el caso de Brasil, donde el dato de 2011 se obtuvo de OCDE (2014) y en Argentina, donde de 2005 en adelante se utilizaron estimaciones propias basadas en CGECSE-MCyE y INDEC-MECON.

Gráfico 96. Evolución del gasto en educación frente al PBI. Países seleccionados. 2000-2014



Fuente: CIPPEC, sobre la base de UIS-UNESCO, excepto para Brasil, cuyo dato de 2011 se obtuvo de OCDE (2014) y Argentina, donde desde 2005 en adelante se utilizaron estimaciones propias basadas en CGECSE-MCyE y INDEC-MECON.

Anexo II El caso de Argentina en la prueba PISA 2015

Introducción general

Dos elementos distintivos de PISA 2015 dificultan la interpretación de los resultados de Argentina: en primer lugar, en la publicación del informe, la OCDE toma la decisión de no incluir a Argentina en la presentación y análisis de los resultados del operativo, basando su argumentación en problemas metodológicos asociados al marco muestral. Los resultados del país quedan relegados a los cuadros del anexo y, junto con Kazajistán y Malasia, se presentan en un recuadro aparte con una advertencia de que la cobertura es demasiado pequeña para garantizar comparabilidad.

En cambio, los resultados de la Ciudad Autónoma de Buenos Aires (CABA) forman parte del cuerpo del informe. Esta decisión se justifica en el hecho de que los problemas metodológicos detectados en la muestra nacional no fueron identificados en la submuestra de CABA.

Es posible incluir los resultados de CABA en forma independiente al resto del país porque desde PISA 2012 es entidad adjudicada de Argentina. Esto significa que el país solicita a la OECD que en cada operativo de PISA se conforme una muestra especial para poder construir estimaciones representativas exclusivamente de la ciudad.

Por otro lado, tanto Argentina como CABA presentan una mejora de sus desempeños muy por encima de los resultados en años anteriores, hecho que complejiza el escenario. La magnitud del incremento tiene escasos antecedentes en PISA.

Como ambos fenómenos (problemas en la muestra y resultados excepcionales) se presentan juntos, favorece a establecer una equivalencia apresurada entre ambos: los desempeños de Argentina 2015 fueron altos porque la muestra tuvo problemas. Esta afirmación es apresurada, ya que no existe ninguna evidencia que permita establecer esta relación con claridad. Por esta razón, es necesario abordar cada uno de estos aspectos por separado. Este anexo se propone introducir algunos elementos que arrojen luz sobre estos aspectos a partir de un trabajo con los resultados de PISA 2015 de acceso público. El alcance de la información disponible permite establecer algunas afirmaciones, y explorar algunas hipótesis.

En el apartado se presentan datos que ponen en evidencia que la muestra de Argentina no cumple con los requisitos mínimos que garantizan la calidad de un relevamiento de esta naturaleza, por lo que en términos estrictamente técnicos es correcto que no se considere comparable.

Complementariamente, se analizan los resultados de PISA 2015 a la luz de algunas hipótesis acerca de las causas por las que el país muestra esa mejora destacada respecto del 2012. Si bien la información estadística es insuficiente para poder arribar a conclusiones, permite poner en juego hipótesis exploratorias.

La exclusión de Argentina en PISA 2015

Para garantizar la calidad de la información que se produce en cada operativo, las acciones del consorcio PISA se planifican y desarrollan en base a un conjunto de estándares técnicos formulados por la OCDE (OECD 2015). Estos estándares conforman criterios ordenadores de las acciones involucradas en las diferentes etapas de la evaluación. Su cumplimiento constituye un compromiso mutuo de las partes, y existen diferentes instancias de monitoreo y seguimiento. En el Anexo IV se presentan los estándares vinculados a la calidad de la información recolectada.

La OCDE tomó la decisión de excluir a la Argentina en la publicación de los resultados de PISA 2015 debido a que en el proceso de recolección de los datos nacionales se incumple con el estándar 1.7, uno de los criterios básicos vinculados a la calidad de la información⁴³.

El estándar 1.7 establece que “La población objetivo definida de PISA (*“Defined Target Population”*)⁴⁴ cubre el 95% o más de la población objetivo deseada de PISA (*“Desired Target Population”*)⁴⁵ (OCDE 2015). La diferencia entre la población objetivo deseada y definida está dada por las exclusiones: escuelas o estudiantes que no se incluyen en el marco muestral (es decir, no pueden ser elegidos para participar de la muestra).

Las exclusiones pueden ser explícitas y programadas. Son admitidas por PISA bajo determinadas condiciones, en acuerdo con cada país, pero no pueden superar el 5% de la población objetivo. Se justifican en función de ciertas características de la población (por ejemplo, estudiantes con ciertos tipos de discapacidad, o minorías lingüísticas) o de las escuelas (por ejemplo, establecimientos muy pequeños, o de muy difícil acceso).

Pero también las exclusiones pueden deberse a errores y omisiones ocurridas en alguna etapa del proceso de muestreo⁴⁶. En estos casos, estas exclusiones no quedan formuladas en forma explícita, pero sus consecuencias son visibles. Una forma de visualizarlas es relacionar la muestra ponderada y la población objetivo: si se manifiestan numerosas exclusiones, la muestra ponderada no alcanza a representar a la población⁴⁷.

Los resultados de Argentina fueron excluidos de la publicación de PISA 2015 debido a que no cumplen este requisito. La población en condiciones de ser seleccionada para participar de la muestra fue menor a la que debiera haber sido. A continuación se presentan algunos datos de cobertura incluidos en el informe PISA:

⁴³ Fuente: Capítulo 14 (*Data Adjudication*) del informe técnico PISA 2015 (2017).

⁴⁴ La población objetivo definida es el conjunto de los estudiantes que, al momento de seleccionar la muestra, poseen una probabilidad estadística de ser evaluados. Es decir, todos los que conforman el marco muestral de PISA.

⁴⁵ La población objetivo deseada es aquella que se propone evaluar PISA: Una muestra representativa de estudiantes de entre 15 años y los 3 meses a 16 años y 2 meses de edad, (considerando como fecha de referencia el inicio del período de exámenes), que asiste a 7mo grado de estudios o más, en instituciones educativas ubicadas dentro del país.

⁴⁶ Cabe señalar en este punto que como “proceso de muestreo” se consideran todos los pasos requeridos por PISA para la selección de la muestra, que quedan plasmados en sus documentos técnicos. Incluye acciones del país que son supervisadas por PISA (como la confección del marco de muestreo) y otras que realiza PISA (como la selección de la muestra de escuelas y estudiantes). Para mayor detalle, ver OECD (2017)

⁴⁷ El Anexo IV desarrolla esta forma de visibilizar problemas de cobertura.

Cuadro 2. Datos de cobertura de PISA 2015 en Argentina

	Argentina
Población total de 15 años	718.635
Matrícula de 15 años en grado 7 o más	578.308
Población objetivo deseada	578.308
Exclusiones de estudiantes a nivel de escuela	2.617
Población objetivo definida	575.691
Muestra ponderada de estudiantes evaluados	394.917

Fuente: CIPPEC, en base a OCDE (2016) y OCDE (2016b)

Los datos presentados en el cuadro permiten identificar que:

- Se estimaba que la población argentina de 15 años era de 718.635 al año 2015, de los cuales 578.308 constituyen población objetivo de PISA, porque asistían a escuelas en séptimo grado en adelante. Todos ellos fueron considerados elegibles para la muestra de PISA 2015.
- Sobre esta población objetivo, se excluyeron en forma explícita 448 escuelas de educación especial, a las que asistían 2.617 estudiantes. Por lo que, en teoría, el marco muestral debía incluir a todas las escuelas a las que asistían los 575.691 estudiantes que conformaban la población objetivo definida, el 99,5% de la población objetivo deseada.
- Sin embargo, la muestra ponderada de estudiantes en Argentina alcanza valores marcadamente menores a los esperados, siendo representativa de 394.917 estudiantes. Esta diferencia refleja el problema de la calidad de los datos: hay una brecha muy marcada entre la población representada y la que se esperaba representar según las definiciones iniciales. La muestra es representativa del 68% de la población objetivo de PISA, esto significa que uno de cada tres estudiantes argentinos de 15 años no está representado en los resultados de PISA.

Estos bajos niveles de cobertura de la muestra de Argentina tienen dos efectos principales:

- El primero y más importante es que los resultados no representan a la población⁴⁸. El principio fundante de la estadística inferencial está en la capacidad de analizar un grupo pequeño de casos (muestra) para realizar afirmaciones sobre un grupo mucho mayor de casos (población). Los resultados de PISA 2015 son reflejo de los desempeños de los 6.349 estudiantes evaluados, pero no es posible afirmar con rigurosidad que estos resultados pueden trasladarse al conjunto de los alumnos argentinos.

⁴⁸ Se expresa de esta manera para facilitar la comprensión del lector. La representatividad refiere a la propiedad de trasladar resultados de una muestra a una población, bajo ciertas condiciones y con márgenes de error calculables. El hecho de que no sean representativos no significa *per se* que sean diferentes (superiores o inferiores), sino que no se han dado las condiciones para trasladar los resultados a la población. Por lo tanto, la afirmación correcta debería ser “no es posible afirmar que los resultados representan a la población objetivo”. La exploración en profundidad de la muestra y su vinculación con datos de la población puede arrojar indicios de la existencia de sesgos en los resultados.

- En segundo lugar, que los resultados no son comparables con aplicaciones anteriores. La comparabilidad requiere como requisito básico poner en relación dos poblaciones similares. En este caso no es posible cumplirlo: no se puede determinar si los cambios en los resultados pueden deberse a diferencias en los desempeños de estudiantes comparables, o si ocurre porque son poblaciones de diferentes características.

Al explorar las razones por la que no se alcanzó la población deseada, el consorcio PISA afirma en el Volumen I del informe PISA 2015 que:

“La muestra de PISA 2015 para Argentina no cubrió la totalidad de la población objetivo, debido a una posible omisión de escuelas en el marco de muestreo, con excepción de la región adjudicada de Ciudad Autónoma de Buenos Aires (Argentina) (en adelante "CABA" [Argentina]); como consecuencia, sólo los resultados para CABA (Argentina) pueden compararse en el tiempo (...).

En Argentina, la evaluación PISA fue realizada de acuerdo con las normas y directrices operativas de la OCDE. Sin embargo, hubo una disminución significativa en la proporción de niños de 15 años cubiertos por la prueba, tanto en cifras absolutas como relativas.” (OCDE 2016)

En el capítulo 14 del reporte técnico de PISA 2015 (OCDE 2016f) se amplía el detalle de las observaciones sobre Argentina, que se resumen en los siguientes puntos:

- El marco muestral presentado por Argentina en febrero de 2015 contenía menos escuelas que en ciclos anteriores. Pero no así en el total de matrícula, donde las cifras de inscripción eran similares.
- Sin embargo, en la etapa siguiente de muestreo en campo (junio-julio 2015), el total real de estudiantes se redujo significativamente respecto del número estipulado en el marco muestral (alrededor del 30%).
- PISA identificó luego del relevamiento que tres categorías de escuelas fueron omitidas en el marco muestral:
 - Escuelas que habían sido creadas o renombradas entre 2013 y 2015, a causa de una política de reestructuración en curso que combinó las escuelas secundarias inferior y superior en nuevas escuelas secundarias unificadas.
 - Escuelas que fueron catalogadas incorrectamente en el registro oficial de instituciones sin estudiantes. Este error fue detectado en octubre de 2015.
 - Finalmente, algunas autoridades provinciales no incluyeron todas las escuelas rurales ya que tenían muy pocos estudiantes.

Para comprender estas afirmaciones, es necesario en primer lugar definir brevemente qué es el marco muestral de PISA. Es el conjunto de escuelas potencialmente elegibles para conformar la muestra donde se aplicará la evaluación. A cada una de las escuelas del marco muestral se le asigna una probabilidad estadística de ser seleccionada, que varía entre escuelas en función de ciertas características específicas de las instituciones y de su población. El marco muestral posee una característica fundamental para PISA: la muestra seleccionada va a ser representativa de las escuelas incluidas en el marco muestral. Por esta razón, los estudiantes inscriptos en las escuelas que forman el marco muestral son considerados la “*población objetivo definida*”, como se ha presentado previamente.

Este concepto es clave: las escuelas que no son incluidas en el marco muestral –y por ende los estudiantes potencialmente elegibles para PISA que asisten a ellas- no serán representadas por los

resultados de PISA 2015. Si esta cantidad de escuelas y estudiantes excluidos supera determinados límites, ya no se puede afirmar que los resultados son representativos de la población objetivo del país.

En el marco muestral de Argentina, según afirma PISA, se excluyeron más escuelas de las que se considera el mínimo necesario para garantizar la representatividad de la muestra. Esto significa incumplir el estándar de calidad técnica de PISA nro. 1.7. Allí se establece que las exclusiones no pueden superar el 5% del total de estudiantes de 15 años que asisten a grado 7 o más.

De esta forma, PISA afirma que “El incumplimiento del estándar 1.7 sobre la definición de la población objetivo y las exclusiones escolares implica que los resultados argentinos pueden no ser representativos de todo el país y, por lo tanto, no ser comparables a los de otros países ni a los resultados de años anteriores” (OCDE 2016b).

Los datos presentados sobre la cobertura de PISA arrojan evidencia clara de que la muestra seleccionada subrepresenta significativamente a la población objetivo, muy por debajo de los límites admitidos.

Sin embargo, las explicaciones otorgadas parecen ser insuficientes para entender cómo fue posible que esa omisión haya pasado los controles de calidad que realiza la OCDE.

Tal como puede observarse, en la hipótesis que propone PISA para explicar la exclusión de escuelas que afectó la representatividad de la muestra, uno de los argumentos principales se centra en que estas omisiones han sido consecuencia de una reestructuración de las escuelas secundarias⁴⁹. Este argumento, vinculado a la unificación de establecimientos con oferta incompleta del nivel, no se desarrolla en profundidad, y llama la atención.

Supone que ciertas instituciones escolares⁵⁰ vieron modificada su estructura, su código administrativo o quizás su localización, en el lapso de tiempo entre la conformación del marco muestral –que suele corresponder a uno o dos años anteriores al ciclo de PISA– y la implementación efectiva de la evaluación en 2015. Frente a esta situación, y asumiendo que existió la capacidad técnica para identificar con anticipación qué escuelas se verían afectadas, el país habría tomado la decisión de excluirlas del marco muestral. ¿Por qué razones no se previeron estrategias para el seguimiento de las instituciones, a conciencia de que este faltante generaría problemas en la representatividad de la muestra?

Considerando que un 31,4% de la matrícula no se ve representada, una estimación simple arrojaría un aproximado de entre 3.500 y 4.000 escuelas excluidas del marco muestral (cantidad proporcional al total de unidades de servicio, sin contar CABA y asumiendo algunos supuestos sobre

⁴⁹ No se analiza la exclusión de escuelas rurales con muy pocos estudiantes, ya que no explicaría el tamaño de la omisión. Tampoco se analiza la existencia de escuelas sin estudiantes en el marco muestral, ya que tampoco es un argumento que permita explicar por qué habría menos escuelas pero igual cantidad de alumnos en el marco muestral. Este error engrosaría el total de escuelas, en lugar de reducirlo.

⁵⁰ En términos técnicos se trata de unidades de servicio, que constituyen los elementos del muestreo de PISA. La unidad de servicio, según el glosario de definiciones para la producción de estadísticas educativas en la Argentina, es la concreción de la oferta de un servicio educativo en una localización (sede o anexo) de un establecimiento. Es decir, toma como referencia la localización geográfica de la oferta, lo que permite contabilizar como unidades diferentes a los anexos, ese decir, localizaciones donde funciona una sección o grupo de secciones que depende pedagógica y administrativamente de una localización “sede” que funciona en otro lugar geográfico (DiNIECE, 2004).

su distribución). Estos valores son marcadamente superiores a lo que las estadísticas oficiales dejan entrever⁵¹.

Tampoco queda claro la razón por la cual el marco muestral no mostraba faltante de matrícula al excluir escuelas. ¿La matrícula de estas escuelas excluidas habría sido “trasladada” o “imputada” a las escuelas que permanecían en el marco muestral? No habría razones para esto.

En síntesis, los datos que presenta PISA son consistentes con una muestra de baja calidad, evidenciando problemas operativos. Las hipótesis que se esgrimen apuntan a errores técnicos en la confección del marco muestral. Sin embargo, las hipótesis que explicarían los motivos de esta exclusión de escuelas (y que a su vez justificarían la detección tardía del problema) son insuficientes y poco claras.

Lo que parece claro es que la muestra de Argentina no alcanza los mínimos estándares de calidad, pero quedan planteadas dudas centrales sobre por qué este problema no pudo ser detectado a tiempo y corregido.

El dilema de los resultados de Argentina en PISA 2015

A este problema de la baja cobertura de la muestra, se le suma otro elemento que genera aún más confusión para comprender el escenario del país: los desempeños de los estudiantes evaluados en 2015 fueron marcadamente superiores a los ciclos anteriores.

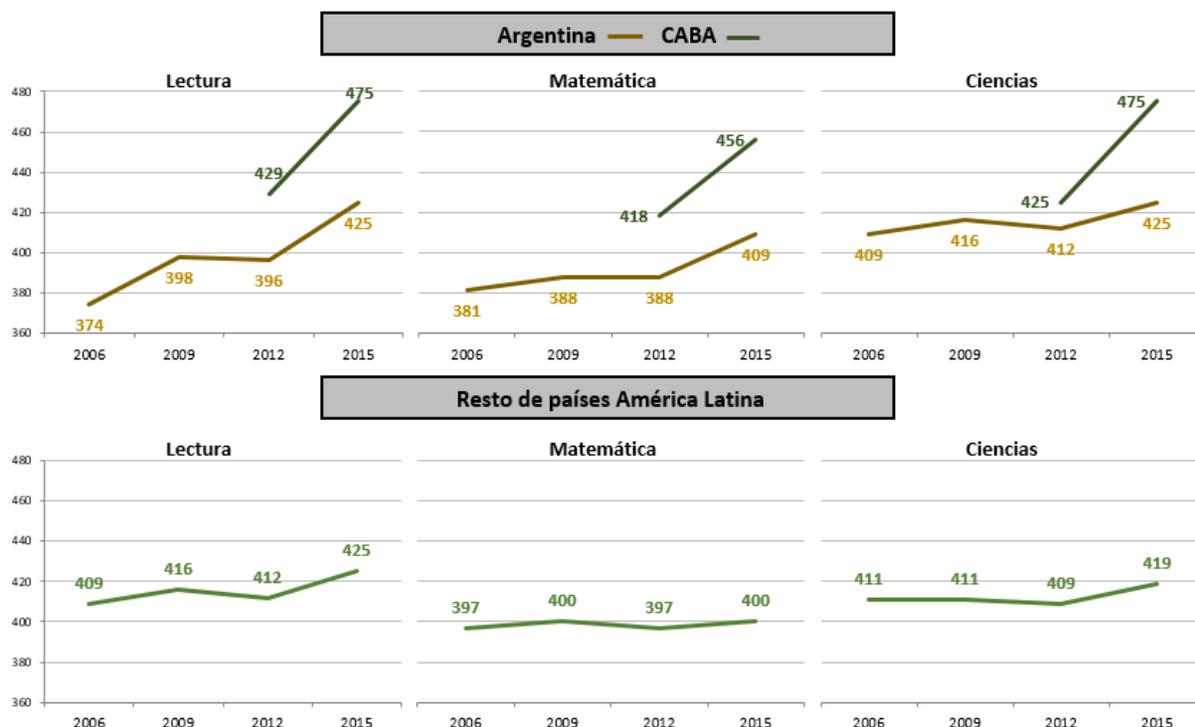
Los resultados han sido tan llamativos, que ha habido una identificación casi automática respecto del problema de la muestra: los resultados de Argentina habrían sido altos porque la muestra no fue de calidad. Sin embargo, esta apreciación es apresurada e incompleta.

Tal como fue explicado en el apartado anterior, las exclusiones del marco muestral derivan en la imposibilidad de afirmar con validez científica que la muestra es representativa de la población. Pero de ello no se desprende que los resultados estén distorsionados o artificialmente incrementados. Para ello, es necesario demostrar que existe algún factor que favorezca la existencia de mejores resultados. Pueden ser de diferente naturaleza, como un sesgo en la muestra hacia población de mayor nivel socioeconómico, o de menor cantidad de repitentes, o distribuidos geográficamente en proporciones distintas a la matrícula, todos elementos asociados a diferentes desempeños.

Por ello, en este Anexo nos proponemos indagar por separado distintos aspectos de la muestra de Argentina en PISA 2015, a los fines de identificar indicios que contribuyan a apoyar diferentes hipótesis sobre los resultados. A continuación, se presentan algunos comentarios sobre los resultados de Argentina en 2015.

⁵¹ Según los datos disponibles, entre los años 2012 y 2015 el porcentaje de unidades de servicio de nivel secundario que ofertaban el nivel completo se incrementó del 71% al 81%, lo que implicó una unificación o ampliación de aproximadamente 1.200 unidades de servicio que ofertaban sólo el ciclo básico, y de unas 200 que sólo ofertaban ciclo orientado. Estos cambios se concentraron en seis provincias (procesamientos propios en base a DiNIEE-MED, Anuario Estadístico Educativo 2012 y 2015)

Gráfico 97. Resultados PISA 2006 - 2015. Argentina, CABA y promedio de países de América Latina (*)



(*) Se consideran los 6 países que, junto con Argentina, forman parte de este estudio.

Fuente: CIPPEC, sobre la base de OCDE PISA 2006, 2009, 2012 y 2015

Argentina muestra para el período 2012 - 2015 una mejora sustantiva de sus resultados en las tres áreas evaluadas, de 26 puntos en promedio. En los tres casos el incremento supera los 20 puntos, en lectura y ciencias se acercan a los 30 puntos. Estas magnitudes representan aproximadamente un cuarto de desvío estándar en la escala de PISA.

CABA sólo participó como región adjudicada en las últimas dos aplicaciones de PISA. Si bien son necesarias al menos tres mediciones para establecer tendencias, los resultados de 2012 - 2015 muestran un incremento de 45 puntos en promedio, con diferencias positivas de 38 puntos en matemática, 46 en lectura y 51 en ciencias.

Estas mejoras en los desempeños no tienen antecedentes en las aplicaciones previas de PISA en el país. Sólo en el área de lectura, entre los años 2006 y 2009, puede identificarse un incremento similar en Argentina (+25 puntos). Pero como no tiene correlato en otras áreas, el promedio global de mejora desciende a 14 puntos. Por lo tanto, esta mejora conjunta de las tres áreas es inédita a nivel nacional.

En los países de la región tampoco se observa una mejora de esa magnitud, tanto en los valores promedio (que se presentan en el gráfico) como en los resultados de cada uno de ellos. Los antecedentes registrados de mejoras sustantivas en las tres áreas pueden identificarse en las aplicaciones 2003 - 2006 de Brasil y Colombia (mejora promedio de 17 puntos en ambos casos), y en las aplicaciones 2012 - 2015 en Perú, Uruguay (mejora promedio de 18 puntos en ambos casos) y Colombia (17 puntos promedio). En ambos casos, la mejora es inferior a la observada en Argentina. Por lo tanto, en los antecedentes de la región, no hay registro de una mejora de esa magnitud en un período de tres años. El caso de Perú, el país de América Latina destacado por mostrar los mayores

niveles de crecimiento en PISA, presenta mejoras de similar magnitud pero en períodos de tiempo más amplios.

Incluso si se consideran todos los países participantes de todas las aplicaciones de PISA (exceptuando el año 2000, por razones de comparabilidad), pueden identificarse sólo cuatro países que han manifestado mejoras superiores a 20 puntos en promedio. Para agregar una complicación adicional al análisis, dos de esos cuatro países (Kazajistán y Malasia 2012 - 2015) comparten una característica con Argentina: sus resultados no son comparables en el año donde se observa la mejora debido a problemas vinculados a la cobertura del relevamiento. Por lo tanto, las mejoras de Argentina y de CABA también son excepcionales en el universo de PISA, considerando todos los años y países participantes.

Mediante esta descripción no se busca introducir sospechas sobre los resultados o socavar su credibilidad (ya en riesgo por los problemas de la muestra), sino contextualizar el desafío de entender los resultados de Argentina en 2015.

A diferencia de otros casos de la región, tampoco es posible relacionar esta mejora en los desempeños con el cambio de metodología de escalamiento de los resultados implementada en 2015, tal como fue presentado en el capítulo 2. Aún con los resultados de Argentina 2012 reescalados, la mejora sería importante: en promedio, las tres áreas incrementan sus resultados en 22 puntos. Lamentablemente no se cuenta con información de los resultados reescalados para CABA.

¿Qué razones pueden explicar esta mejora en los desempeños? ¿Existe una relación con el problema del marco muestral? ¿Qué resultados hubiese obtenido el país si no se hubiesen manifestado problemas en el marco muestral?

Es necesario identificar hipótesis razonables que permitan esbozar explicaciones, y que puedan tener consistencia con el comportamiento de los datos. Se propone a continuación explorar en profundidad los datos públicos disponibles de PISA 2015 para identificar algunas señales que contribuyan a la interpretación de estos resultados.

Análisis de la composición de la muestra de Argentina

Un primer aspecto a explorar es si ha habido cambios en la composición interna de la muestra de PISA 2015 respecto de 2012. Considerando los problemas detectados es posible que haya variaciones en la población evaluada que incidan en los resultados.

Una hipótesis a contrastar se centra en la posibilidad de que las escuelas omitidas en el marco muestral hayan sido excluidas intencionalmente por tratarse de establecimientos que tendrían bajo desempeño, como instancia de manipulación. Una hipótesis alternativa puede establecer, sin necesidad de afirmar una intencionalidad, que por razones vinculadas al azar, la muestra seleccionada sobre un conjunto reducido de escuelas arrastre ciertos sesgos que incidan en los resultados. En ambos casos, cambios en las características de la población evaluada pueden explicar modificaciones llamativas en los resultados.

La investigación en educación ha dado más que abundantes evidencias de la asociación estadística entre ciertas características del estudiante, de su contexto o de la institución escolar con los desempeños en las evaluaciones. Por ejemplo, el nivel educativo del hogar, la acumulación de experiencias de repitencia, o la asistencia a escuelas de gestión estatal o privada, suelen arrojar marcadas diferencias. En general, las muestras bien confeccionadas controlan estas características para garantizar poblaciones comparables.

Si existieran en PISA 2015 poblaciones diferentes a las evaluadas en ediciones anteriores (lo que podría ser consecuencia de los problemas en la cobertura de la muestra de Argentina)⁵², con mayor presencia de características asociadas a mejores resultados, se podría explicar el incremento de los desempeños. Partiendo de este supuesto, se presentan y analizan a continuación algunos datos e indicadores de la muestra ponderada⁵³ de Argentina en PISA 2012 y 2015, para identificar variaciones.

Cuadro 3. Datos e indicadores seleccionados de PISA 2012 y 2015. Argentina.

		TOTAL ARGENTINA		
		PISA 2012	PISA 2015	Diferencia 2012-2015
Muestra	<i>Total de casos</i>	5.908	6.349	+441
	<i>Población representada (muestra ponderada)</i>	545.941	394.917	-151.024
Sexo	<i>Mujeres</i>	51,0%	52,0%	+1,0%
	<i>Varones</i>	49,0%	48,0%	-1,0%
Grado al que asiste	7	2,0%	1,0%	-1,0%
	8	12,0%	9,0%	-3,0%
	9	23,0%	23,0%	+0,0%
	10	59,0%	64,0%	+5,0%
	11	3,0%	3,0%	0,0%
	12	1,0%	0,0%	-1,0%
Índice ESCS	<i>Media</i>	-0,70	-0,79	-0,09
	<i>Percentil 25</i>	-1,49	-1,64	-0,15
	<i>Percentil 75</i>	0,14	0,12	-0,02
Experiencia acumulada de repetencia	<i>Nunca repitió</i>	63,7%	71,1%	+7,4%
	<i>Repitió una vez</i>	21,9%	21,7%	-0,2%
	<i>Repitió dos veces o más</i>	14,4%	7,2%	-7,2%
Nivel Educativo de la madre	<i>Secundaria orientada</i>	53,7%	50,2%	-3,5%
	<i>Secundaria básica</i>	20,0%	18,4%	-1,6%
	<i>Primaria</i>	19,0%	23,7%	+4,7%
	<i>No terminó la primaria</i>	7,3%	7,7%	+0,4%
Sector de gestión	<i>Público</i>	66,4%	78,5%	+12,1%
	<i>Privado</i>	33,6%	21,5%	-12,1%

Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

⁵² También es importante considerar, tal como se desarrolló en el capítulo 2, que pueden existir problemas de comparabilidad independientes del problema de exclusión de escuelas que afecta a Argentina. Si no se controlan correctamente los factores potenciales de sesgo en el diseño de la muestra, se pueden construir estimaciones con problemas de comparabilidad. A veces estos factores no son sencillos de identificar y prever.

⁵³ La ponderación de la muestra refiere a la realización de cálculos en base a una muestra en los que se considera la cantidad de casos de la población que representa cada unidad de muestra. De esta forma, el uso de ponderadores permite estimar los totales de población, y también se utilizan para ajustar la muestra cuando el método de muestreo, u otras características del proceso de construcción de la muestra (como la no respuesta), provoca que ciertas características estén representadas en proporciones diferentes a la población. En otras palabras, los resultados de una muestra ponderada son los estimadores de las características de la población

El cuadro muestra algunas tendencias relevantes de señalar:

- a) El tamaño de la muestra de estudiantes en 2015 fue mayor a la del 2012, aunque no tiene ninguna implicancia para los resultados. La población representada evidencia una significativa caída, expresión del problema de exclusión en el marco muestral.
- b) En el año 2015 hay una mayor presencia de estudiantes en grado 10 (+5,0%), que es el grado que corresponde teóricamente a los 15 años de edad, y menor presencia en grados menores, lo que implicaría porcentajes más bajos de sobreedad en la muestra.
- c) Consecuentemente, hay mayor presencia de estudiantes que declaran no haber repetido nunca (+7,4%) por sobre los que repitieron dos o más veces (-7,2%)
- d) Los hogares presentan un clima escolar más bajo en 2015, con mayor presencia de madres con primaria completa (+4,7%) y menos con secundaria completa (-3,5%).
- e) Asimismo, el indicador de PISA de contexto socioeconómico (ESCS) muestra una media levemente más baja en 2015, con mayor brecha si se considera al percentil 25%. Es decir, la muestra del 2015 abarcaría a una población con mayores niveles de pobreza, aunque con diferencias leves.
- f) Hay una marcada participación mayor de estudiantes de escuelas de gestión estatal (+12,1%).

Las diferencias identificadas en los puntos (b) y (c) podrían contribuir a suponer resultados más altos en 2015, al haber menos repitentes. Pero lo llamativo de estos resultados es que los cambios en la composición de la muestra ponderada señalados en (d), (e) y (f) son de sentido contrario: hay en 2015 una población con mayores niveles de pobreza, menos nivel educativo en el hogar, y mayor participación del sector estatal, todas características asociadas a más bajos resultados.

Algunos de estos cambios observados pueden ser contrastados con fuentes externas, a los fines de reconocer si se asocian a modificaciones en las trayectorias o en la composición socioeducativa de la población.

En primer lugar, las estadísticas oficiales de educación en Argentina reflejan cambios en el indicador de sobreedad de los estudiantes de 15 años relativamente consistentes con estos valores. Se considera para este análisis el indicador de sobreedad, ya que la pregunta de PISA indaga sobre experiencias de repitencia en la trayectoria escolar previa.

Entre 2012 y 2015, el porcentaje de estudiantes de 15 años que asisten a 7mo grado o más con sobreedad se redujo del 44,1% al 39,1% (procesamientos propios en base a datos de DiNIEE-MED, Relevamiento Anual de Matrícula y Cargos). Aunque la magnitud de la diferencia en la sobreedad (-5,0 puntos) es menor a la observada en la declaración de repitencia (-7,4 puntos), podría establecerse cierta relación en esta reducción respecto de la existencia de trayectorias más oportunas.

Por otra parte, para analizar las variaciones en el nivel educativo de la madre, se realizó un procesamiento sobre la Encuesta Permanente de Hogares de Argentina (EPH). Para construir una población equivalente, se identificó a la población de mujeres jefas de hogar o cónyuges de entre 35 y 55 años en hogares donde habitaban adolescentes de 15 años que asisten a la escuela, tanto en la EPH 2012 como 2015 (segundo trimestre en ambos casos).

Como resultado, se observó en este grupo de mujeres una reducción del porcentaje con máximo nivel educativo alcanzado primaria completa o menos, del 37% en 2012 al 30% en 2015. Es decir, una reducción de 7 puntos porcentuales, lo que constituye una tendencia opuesta a las diferencias

observadas⁵⁴. Aquí podrían presentarse sesgos en la muestra de 2015, pero nuevamente implicarían un sesgo hacia desempeños más bajos, ya que se evaluaría una población más vulnerable.

Por último, el porcentaje de matrícula de 15 años en grado 7 o más que asiste a educación de gestión estatal ha permanecido estable en estos años, variando del 72,1% (2012) al 71,1% (2015)⁵⁵. En este caso, llama la atención no sólo la variabilidad de estos factores, sino también el hecho de que ninguna de las dos muestras arroje porcentajes similares a la población. El valor real queda situado en el punto medio entre las estimaciones 2012 y 2015.

De estas observaciones se puede deducir que no hay características de la muestra, al menos en las variables clave que se pueden analizar desde PISA, que estén claramente asociadas a predecir mayores desempeños en 2015.

Este análisis permite afirmar que las poblaciones que han sido evaluadas en PISA 2012 y 2015 no muestran señales de sesgos que permitan explicar mejores resultados en la última aplicación. Ambas son similares en algunas características. Hay variaciones observadas que se relacionan con cambios en la población objetivo. Otras son en sentido contrario a lo que cabría suponer: mayor presencia en 2015 de factores asociados a bajos logros.

Análisis de distribución de resultados por escuela

En este escenario donde la Argentina – y también la Ciudad de Buenos Aires – ha obtenido desempeños en PISA 2015 llamativamente superiores a 2012, es ineludible situar la mirada sobre las hipótesis que podrían explicar estos resultados, ya sea que estos resultados expresen una mejora de la calidad de los aprendizajes en el país, o sean expresión de algún factor que esté introduciendo sesgos. Si bien la información disponible no permite elaborar afirmaciones concluyentes, es posible reconocer comportamientos consistentes con ciertas explicaciones.

En este marco, mirar los resultados por escuela es otra vía de análisis que puede arrojar pistas para entender el comportamiento de los resultados. Las bases de datos disponibles permiten reconstruir los desempeños del conjunto de estudiantes de cada institución participante (sin posibilidad de identificarla). De esta forma, se pueden mirar los resultados de cada escuela y compararlos con el comportamiento de los resultados en 2012. Si bien la muestra de escuelas no es la misma, es esperable que se identifique un reparto más o menos equivalente entre escuelas con mayores y menores desempeños.

Este ejercicio parte del supuesto de que la muestra del año 2015 es comparable con la del 2012 porque, como se analizó previamente, muestran una composición similar. Cabe señalar que este supuesto es independiente del problema de cobertura: en este ejercicio no se analizan los resultados proyectados a la población, sino las muestras seleccionadas en cada operativo.

Esta exploración se realiza con la intención de analizar qué diferencia a la distribución de escuelas en 2015 respecto de 2012. Se busca identificar algunas características que podrían asociarse a diferentes hipótesis, por ejemplo:

- a) Presencia de un mayor número de escuelas con resultados llamativamente destacados

⁵⁴ Procesamientos propios en base a INDEC, microdatos de la Encuesta Permanente de Hogares, años 2012 y 2015 – segundo trimestre.

⁵⁵ Procesamientos propios en base a datos de DiNIEE-MED, Relevamiento Anual de Matrícula y Cargos.

- b) Resultados poco consistentes con el contexto socioeconómico de las escuelas
- c) Mejora generalizada del conjunto de escuelas, con similares proporciones, y resultados que mantienen la forma de curva normal
- d) Distribución diferente de respuestas correctas entre diferentes tipos de ítems
- e) Reducción de escuelas con muy bajos resultados
- f) Reducción de escuelas con porcentajes altos de no respuesta e ítems NRI (ver capítulo 2)

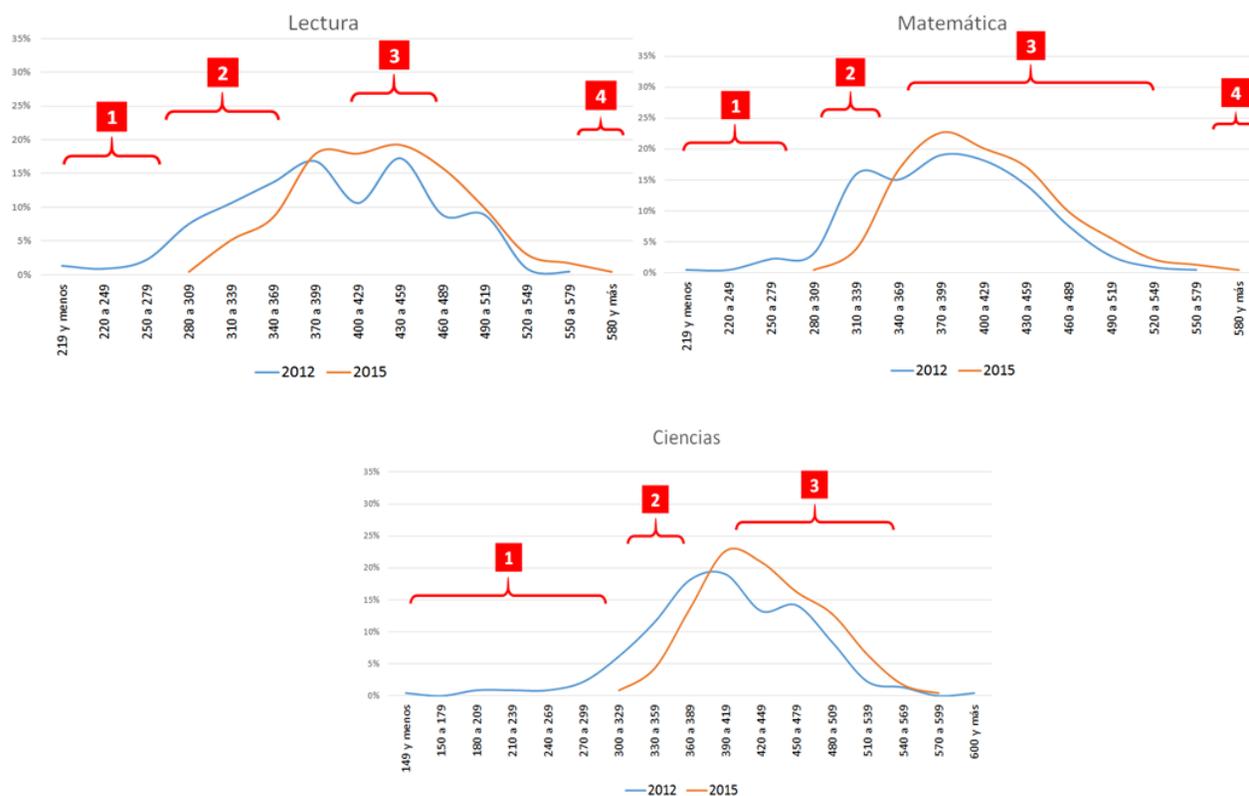
La existencia o ausencia de alguna/s de esta/s características no constituye *per se* evidencia concluyente. En el marco de posibles hipótesis que permitan explicar las razones de este crecimiento en los desempeños entre 2012 y 2015, se propone identificar datos *consistentes* con algunas de ellas. Se mencionan algunos ejemplos posibles:

- La identificación de las situaciones a y b podría abonar las hipótesis vinculadas a una intervención externa que haya facilitado las respuestas o “ayudado” a los estudiantes a responder.
- La situación c sería consistente con la hipótesis de que el incremento en los desempeños se asocia a una mejora en los aprendizajes: las intervenciones externas tienden a distorsionar las distribuciones.
- En cambio, reconocer situaciones como la e y la f favorecerían el planteo de hipótesis asociadas al impacto positivo de acciones de “sensibilización”, que inciden en una menor tasa de ítems no respondidos.
- Otro ejemplo posible es el caso d. Si se detectaran cambios en el porcentaje de respuestas correctas en los ítems de respuesta abierta, podría asociarse a la hipótesis de que la mejora se explica por la implementación de criterios más laxos de corrección.

En este apartado se presentan algunos procesamientos que permiten analizar estos y otros comportamientos, sobre la base de la distribución de resultados por escuela.

Los gráficos que se presentan a continuación muestran cómo se distribuyen las escuelas según sus resultados en 2012 y 2015, agrupados según rangos.

Gráfico 98. Porcentaje de escuelas según rango de resultados. PISA 2012 y 2015. Argentina



Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

A través de la comparación de las líneas de distribución de escuelas en ambos años, es posible perfilar cómo queda expresada la mejora generalizada de resultados en el conjunto de escuelas de la muestra.

Las tres áreas muestran un perfil con estructura coincidente. Se han identificado en cada caso cuatro tramos diferentes:

1 Rangos de resultados muy bajos donde sólo se ubican escuelas evaluadas en 2012. En general este tramo abarca tres categorías de rangos, a excepción del área de ciencias donde llega a abarcar los 6 rangos más bajos de resultados (120 a 299 puntos). En estos rangos se ubica entre el 3% y el 5% de las escuelas evaluadas en 2012.

En el año 2015 ninguna escuela obtuvo resultados muy bajos.

2 Rangos de resultados bajos, donde hay mayor concentración de escuelas en 2012 respecto de 2015. Estos tramos pueden ser de dos a tres rangos de puntaje. Abarca aproximadamente a un 20% (Matemática) o 30% (Lectura, Ciencias) de las escuelas evaluadas en 2012, y a un 5% (Matemática) a 12% (Lectura, Ciencias) de las escuelas evaluadas en 2015

En el año 2015 el porcentaje de escuelas con resultados bajos se redujo a menos de la mitad.

3 Rangos de resultados medios - altos, donde hay mayor concentración de escuelas en 2015 respecto de 2012. Este tramo es extenso en Matemática (6 rangos de resultados) y Ciencias (5 rangos), y menor en Lengua. En 2015, abarca aproximadamente al 90% de las escuelas

evaluadas en Matemática (77% en 2012), al 80% en Ciencias (57% en 2012) y al 50% en Lengua (37% en 2012)

En el año 2015 el porcentaje de escuelas con resultados medios y altos se incrementó

4 Rangos de resultados muy altos, donde no hay escuelas en 2012 pero sí en 2015. Es decir, escuelas con resultados muy superiores en 2015. Este tramo abarca sólo un rango en Lectura y Matemática, y no se identifica en ciencias

El año 2015 solo una escuela alcanzó resultados superiores al mayor puntaje obtenido en 2012.

¿Qué permiten identificar estos resultados? Que la mejora de los puntajes en 2015 se explica principalmente por:

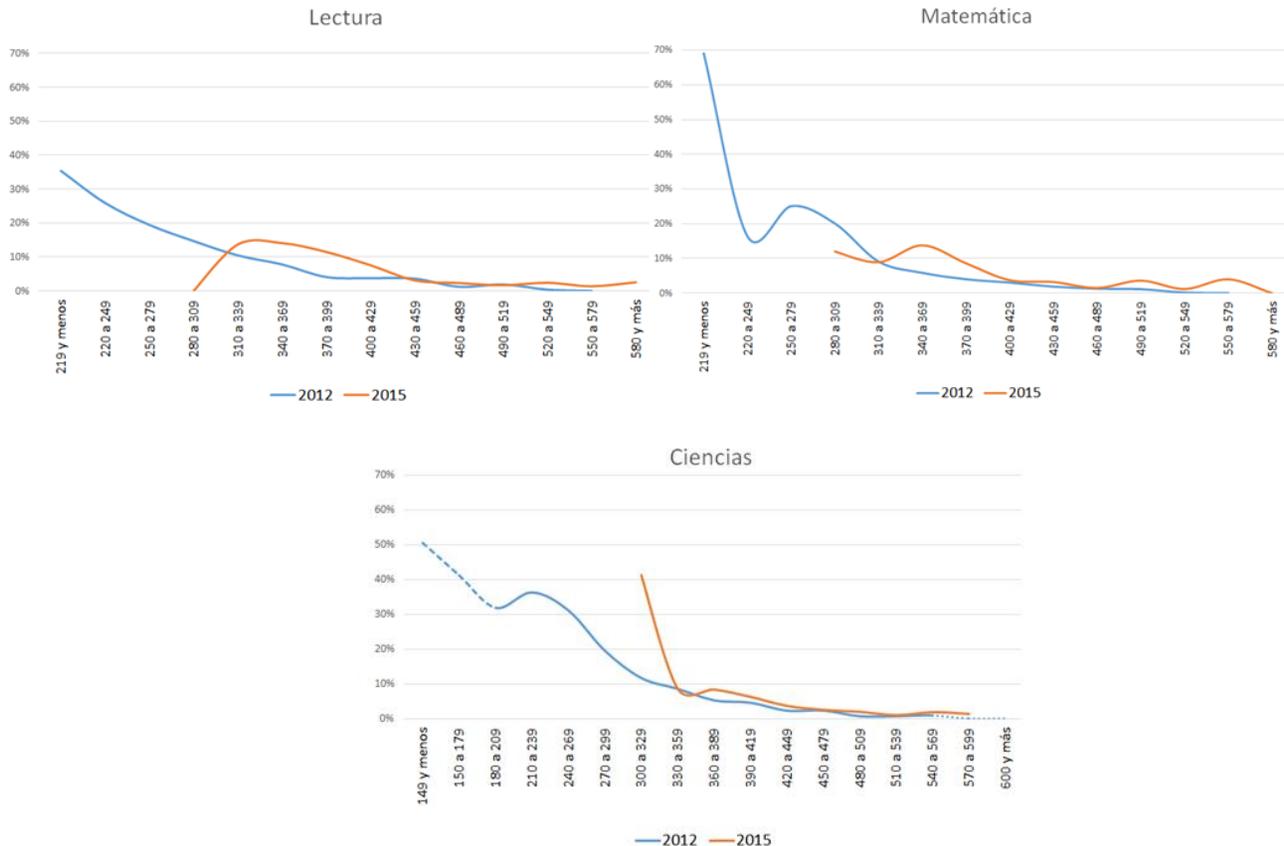
- a. En 2015, ninguna escuela obtuvo resultados muy bajos
- b. Decrece el porcentaje de escuelas con resultados bajos
- c. Se incrementa el porcentaje de escuelas con resultados medios y altos
- d. No se incrementó significativamente la cantidad de escuelas con resultados muy por encima de la media.

En una primera mirada, la distribución de resultados por escuela asume un perfil similar al de la curva normal. Este comportamiento no es consistente con sesgos típicos que podrían reconocerse en una intervención externa, como podría ser una concentración atípica de escuelas con resultados muy por encima de la media. Es necesario seguir explorando los datos para perfilar mejor el comportamiento de esas escuelas.

Para ello, se propone a continuación analizar el perfil de las respuestas a la prueba que presentan esas escuelas con resultados muy bajos en 2012, y compararlas con las respuestas del 2015.

En primer lugar, se analiza el porcentaje de *non-reached items* (ítems dejados en blanco al final de la prueba). Se presenta el dato para cada rango de resultados representado en los gráficos anteriores.

Gráfico 99. Porcentaje de *non-reached items* sobre el total de ítems aplicados, en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina



Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

Como puede observarse en el gráfico, uno de los aspectos que caracteriza a las escuelas con muy bajos resultados en 2012 es el alto porcentaje de *non-reached items*, que llega a duplicar, e incluso triplicar, a las escuelas que obtuvieron resultados superiores.

Por ejemplo, en el área de Lectura, en las escuelas evaluados en 2012 con menos de 280 puntos, en promedio los estudiantes han dejado sin completar al final de la prueba uno de cada cuatro ítems aplicados (26,8% del total). Este porcentaje desciende al 5% en el resto de las escuelas participantes en 2012, y se sitúa en torno al 5,9% en el conjunto de escuelas evaluadas en 2015.

En Matemática 2012, en las escuelas con menos de 280 puntos, en promedio los estudiantes han dejado sin completar al final de la prueba uno de cada tres ítems aplicados (32,0%). En algunas escuelas en particular, en promedio los estudiantes han dejado más de la mitad de los ítems al final de la prueba sin responder.

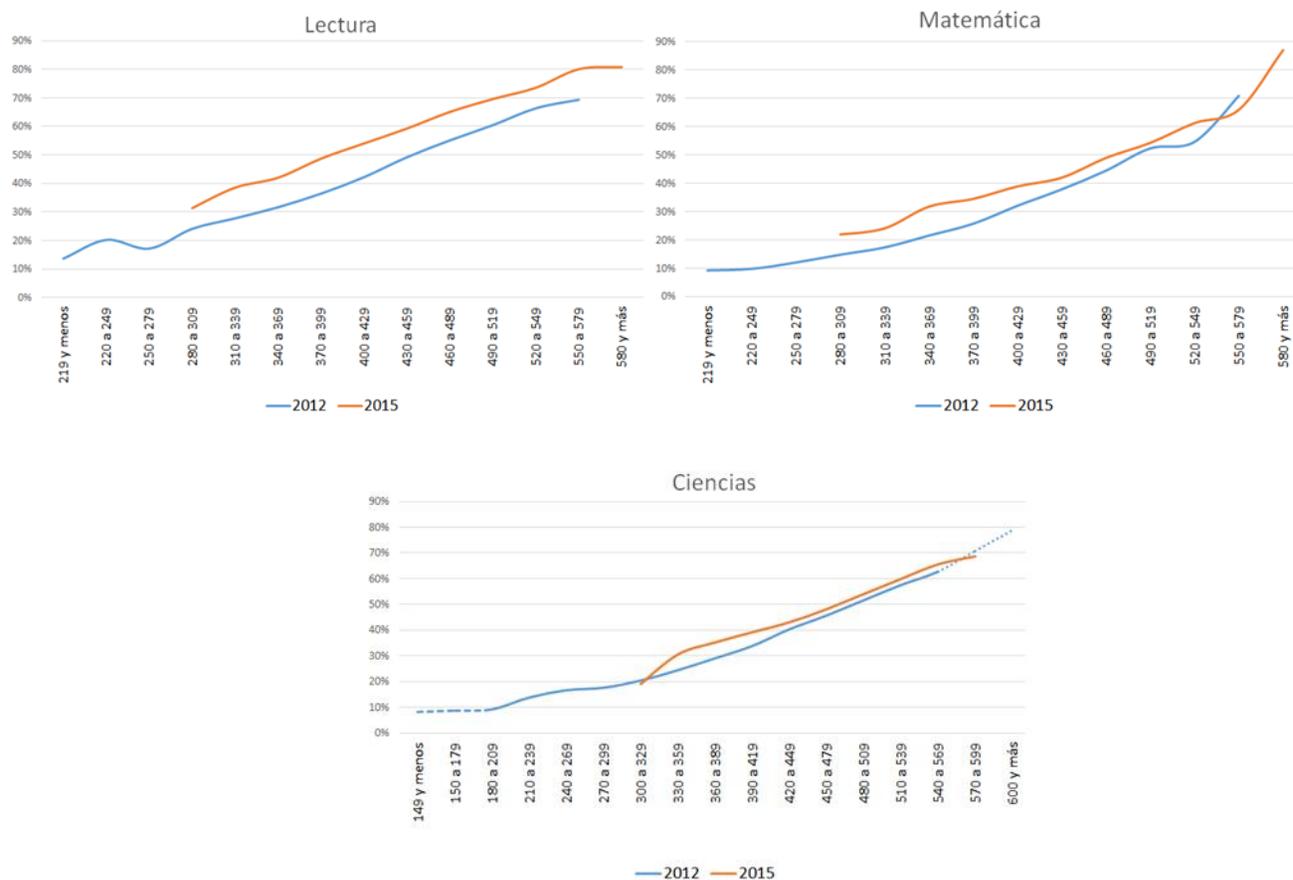
Es importante resaltar la diferencia entre responder un ítem en forma correcta o incorrecta, no responderlo, o dejar una evaluación incompleta. Los *non-reached items* refieren específicamente a este último caso. Representan a estudiantes que “abandonan” la prueba antes de finalizarla o se quedan sin tiempo para completarla. Un análisis en detalle de estos ítems se aborda en el capítulo 2.

Estos datos ponen de manifiesto que en el año 2012, en algunas escuelas fue generalizado el comportamiento de no completar la prueba. Las escuelas con muy bajos resultados fueron aquellas

donde muchos estudiantes dejaban la prueba incompleta. En 2015 esto se revierte: ha habido una menor proporción de estudiantes que no respondieron a las consignas, independientemente que lo hayan hecho en forma correcta o incorrecta.

Este comportamiento de las escuelas con muy bajos resultados es consistente con escenarios donde la aplicación de la prueba es muy poco valorada o las escuelas no están preparadas para recibir a PISA: los estudiantes y docentes restan importancia al acto de evaluación, y completan sus formularios de manera parcial e incompleta.⁵⁶ En forma complementaria, se analiza a continuación el porcentaje de respuestas correctas obtenido por cada grupo de estudiantes. El porcentaje se calcula sobre las respuestas efectivas.

Gráfico 100. Porcentaje de respuestas correctas sobre el total de ítems válidos (descontando no respuesta y NRI), en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina



Nota: se excluyen del cálculo los ítems que tienen crédito parcial, ya que no fue posible acceder al criterio de ponderación de estos casos para identificar la respuesta correcta. Se asume que esta exclusión no afecta las

⁵⁶ Jodouin (2014) analizando los factores que inciden en las tasas de no respuesta en los diferentes contextos culturales y lingüísticos de Canadá, reconoce comportamientos diferentes de los estudiantes frente a la opción por no responder un ítem, arriesgar una respuesta, o abandonar la prueba antes de finalizarla. Estos comportamientos se vinculan con diferencias en lo que el autor llama "cultura de la evaluación". Es decir, en qué medida en la escuela donde se aplica la evaluación existe una comprensión y valoración de sus objetivos por parte de docentes y estudiantes, y existe un marco de referencia claro y compartido para que los estudiantes respondan con esfuerzo.

tendencias totales teniendo en cuenta que son pocos los ítems no considerados. Para el 2012 se excluyeron: 1 ítem de 44 en Lectura, 9 ítems de 109 en Matemática, 3 ítems de 53 en Ciencias. Para el 2015: 8 ítems de 88 en Lectura, 7 ítems de 71 en Matemática, 3 ítems de 90 en Ciencias
Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

Puede observarse como escenario general una relación lineal entre resultados y porcentaje de respuestas correctas en las aplicaciones en ambos años, lo que resulta lógico.

Se observa que las curvas son muy similares entre ambos años, lo que es una observación interesante: las escuelas con mejores resultados se comportan de manera similar en las dos aplicaciones, no se identifican en 2015 comportamientos fuera de la tendencia, como podría ser una marcada concentración de respuestas correctas.

Por otro lado, se observa que uno de los aspectos que caracteriza a las escuelas con muy bajos resultados en 2012 es el escaso porcentaje de respuestas correctas. Cabe recordar aquí que el cálculo considera sólo a los ítems con respuestas efectivas. Es decir, las pruebas en estas escuelas tuvieron muchos ítems sin completar al final de la prueba (*non-reached items*), y además los estudiantes han equivocado la gran mayoría de las respuestas dadas. En promedio, los estudiantes de estas escuelas erraron entre 8 y 9 respuestas de cada 10.

Para alimentar la interpretación de estos resultados, resulta interesante resaltar el efecto del tipo de ítem. PISA incluye en sus evaluaciones ítems cerrados y abiertos. Los ítems cerrados son de tipo *multiple choice*, que pueden ser en formato clásico de preguntas con una opción de respuesta, o en formato complejo (*complex multiple choice*) que usualmente ofrece opciones del tipo Verdadero/Falso en relación a ciertos escenarios o estados. Los ítems abiertos requieren la intervención de un corrector. A grandes rasgos se dividen en aquellos donde la corrección es sencilla y no requiere interpretación (*Constructed Response Manual*) de aquellos que involucran criterios de interpretación, y para lo cual los correctores deben estar entrenados.

Cuadro 4. Porcentaje de ítems NRI y % de respuestas correctas sobre el total de ítems válidos (descontando no respuesta y NRI), por tipo de ítem. Total escuelas 2012 y 2015, y escuelas con muy bajos resultados en 2012. Argentina

Tipo de ítem	PISA 2012				PISA 2015	
	% Ítems NRI		% respuestas correctas (*)		% Ítems NRI	% respuestas correctas (*)
	Total	Escuelas con muy bajos resultados	Total	Escuelas con muy bajos resultados		
<i>Complex Multiple Choice</i>	5,2%	23,7%	30,7%	13,3%	3,6%	42,3%
<i>Simple Multiple Choice</i>	5,3%	23,2%	45,9%	26,1%	4,2%	56,9%
<i>Constructed Response Expert</i>	5,4%	22,6%	26,4%	12,1%	6,1%	43,4%
<i>Constructed Response Manual</i>	5,0%	22,7%	39,1%	20,9%	4,9%	42,7%

(*) Se excluyen del cálculo los ítems que tienen crédito parcial, ya que no fue posible acceder al criterio de ponderación de estos casos para identificar la respuesta correcta. Se asume que esta exclusión no afecta las tendencias totales teniendo en cuenta que son pocos los ítems no considerados. En todos los casos se trata de ítems del tipo Constructed Response

Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

Puede observarse que tanto en términos del % de respuestas correctas, como en el porcentaje de *non-reached items*, se presentan variaciones respecto del tipo de ítem aplicado. En general, las menores dificultades se observan en los ítems *multiple choice* simples en ambos años. En 2012, las dificultades parecieran concentrarse en los ítems abiertos que requieren corrección de experto. Esta diferencia no se visualiza en 2015, donde el porcentaje de respuestas correctas es similar a otros tipos de ítems.

Las respuestas correctas se incrementan en 2015 respecto de 2012 en todos los tipos de ítems: no hay evidencias claras de que exista una relación entre mejores respuestas en un tipo particular de ítem y el incremento de los resultados de 2015. La imposibilidad de acceder a los modelos de evaluación limita avanzar en la exploración de mayor profundidad sobre el comportamiento de las diferentes consignas.

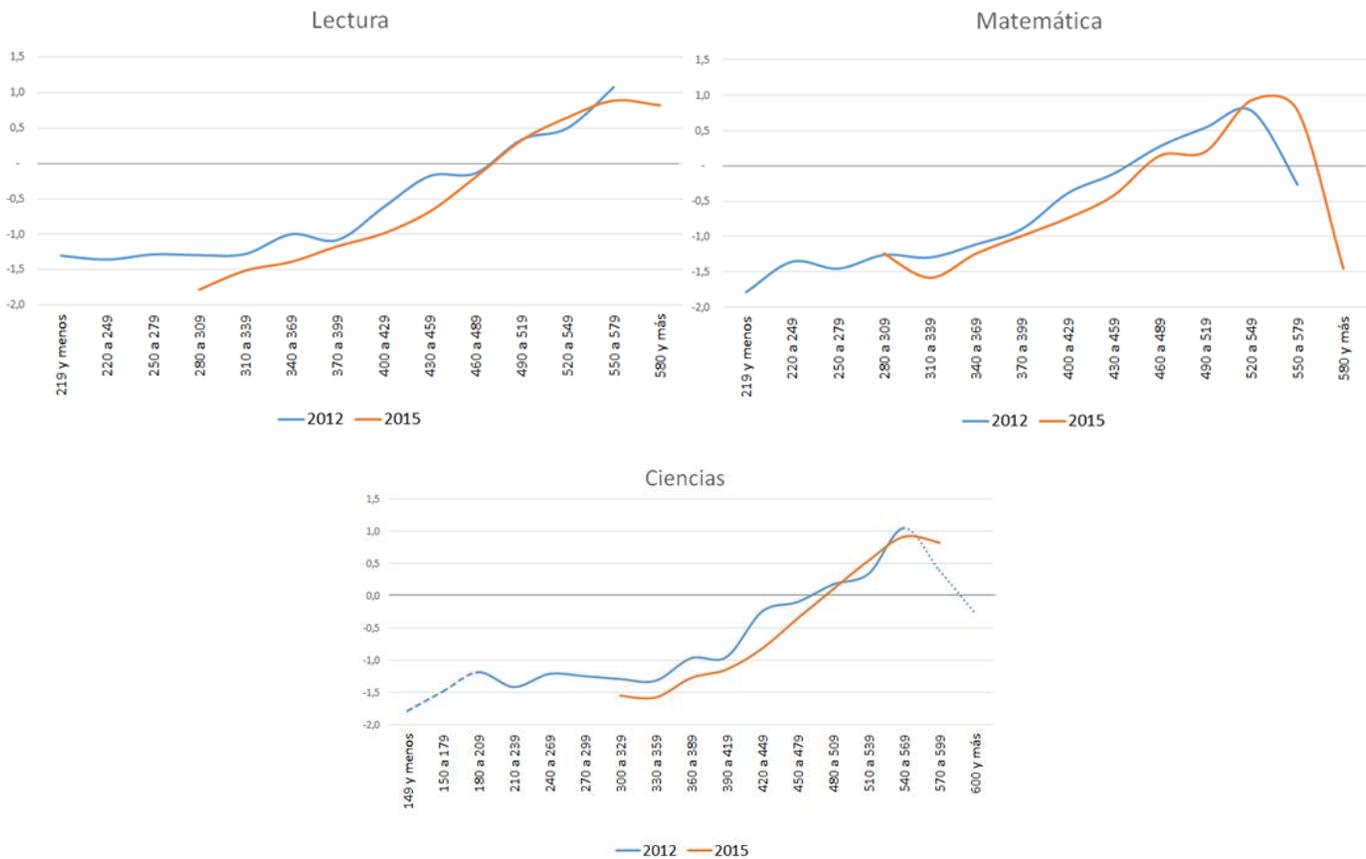
Respecto de las escuelas que muestran mayores dificultades, los porcentajes de ítems NRI son superiores al promedio, y los porcentajes de respuesta inferiores, pero con un comportamiento similar entre diferentes tipos de ítem al promedio general. No parece manifestarse algún comportamiento específico asociado al tipo de ítem⁵⁷.

⁵⁷ Resulta interesante mencionar aquí el caso de Kazajistán. Los resultados del país no se consideran representativos en PISA 2015. Según la descripción que se presenta en el capítulo 14 del reporte técnico, PISA detecta en este país una tasa de respuestas correctas muy superior a la esperada en los ítems que implican corrección manual. El consorcio recodificó una muestra de 300 pruebas y llegó a la conclusión de que los correctores nacionales habían sido muy “indulgentes” en sus

Un aspecto particular de estos resultados llama la atención: aproximadamente la mitad de las consignas de la evaluación de PISA son ítems del tipo *multiple choice*, generalmente con 4 o 5 opciones de respuesta. Un estudiante que responda completamente al azar tiene en estos ítems una probabilidad de acertar la respuesta correcta entre el 20% y 25%. En las escuelas con muy bajos resultados, los porcentajes de respuestas correctas en los ítems de *multiple choice* simple asumen valores levemente por encima de la distribución explicada por azar.

Otro aspecto interesante a indagar es el nivel socioeconómico de los estudiantes que asisten a estas escuelas con muy bajos resultados. En el gráfico a continuación se presenta el índice que construye PISA para caracterizar el nivel socioeconómico y cultural del hogar (índice ESCS, por sus siglas en inglés).

Gráfico 101. Promedio índice nivel socioeconómico y cultural (ESCS) en grupos de escuelas clasificados según rango de resultados. PISA 2012 y 2015. Argentina



Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

critérios. Este caso pone de manifiesto que PISA implementa controles de consistencia sobre los ítems que tienen intervención humana en la corrección(OCDE 2016f)

Cabe recordar que las tendencias globales del indicador entre años: los valores del índice ESCS comparable en el tiempo era de -0,70 en 2012 y -0,79 en 2015 (ver tabla *). Esto reviste vital importancia, ya que confirma que las escuelas evaluadas en ambos años atienden a población de similar nivel socioeconómico.

De la observación de los gráficos podemos obtener algunas apreciaciones interesantes. En primer lugar, hay una relación bastante clara entre el nivel socioeconómico promedio de la escuela y su resultado en PISA.

Sin considerar los extremos (que serán analizados a continuación), el ángulo de la pendiente es muy similar entre ambos años. Esto representa que la incidencia del nivel socioeconómico en los resultados no se modificó sustantivamente entre ambas aplicaciones.

Efectivamente, la relación de resultados y nivel socioeconómico es, a nivel de escuela, más fuerte en el 2015 que en el 2012, particularmente en las áreas de Lectura y Ciencias. El coeficiente de determinación de la relación de ambas variables (R^2) permite identificar este hecho: para el año 2012, el R^2 es de 0,558 en Lectura y 0,508 en Ciencias. En 2015 los valores son 0,593 y 0,584 respectivamente. Esto es sumamente importante, ya que la existencia de intervenciones ajenas a la evaluación que provoquen sesgos positivos en los resultados debería distorsionar esta relación entre resultados y contexto.

En el extremo inferior del rango se identifican dos observaciones a su vez interesantes: las escuelas con muy bajos resultados en 2012 son las más pobres de la distribución. Pero en este grupo se interrumpe la pendiente que marca la relación entre nivel socioeconómico y resultados. En matemática se suaviza, y en lectura y ciencias se mantiene relativamente estable. Por ejemplo, el grupo de escuelas que obtuvo entre 280 y 309 puntos en lectura muestra una composición socioeconómica similar a las que obtuvieron menos de 250 puntos. Cabe preguntarse por qué en este grupo de escuelas con muy bajos resultados el nivel socioeconómico parece no incidir para explicar las diferencias entre escuelas.

En general, la curva del año 2015 tiende a ubicarse levemente por debajo de la del 2012, particularmente en las categorías intermedias y bajas: esto indica mayor presencia de escuelas más pobres en estas categorías. Esta diferencia es razonable, ya que – a igual nivel socioeconómico entre ambas muestras – las escuelas más pobres tuvieron mejores desempeños en 2015.

Llama la atención también que en los valores extremos superiores tiende a manifestarse una disminución el índice, particularmente en Matemática (2012 y 2015) y Ciencias (2012). Como estas categorías superiores suelen estar representando sólo a una escuela, es recomendable omitir estos valores.

Los análisis propuestos en las páginas anteriores permitieron identificar que uno de los principales motivos que explica la mejora en los resultados de Argentina es el hecho de que un conjunto de escuelas con características similares en ambos relevamientos tuvieron mejores resultados en 2015 que en 2012.

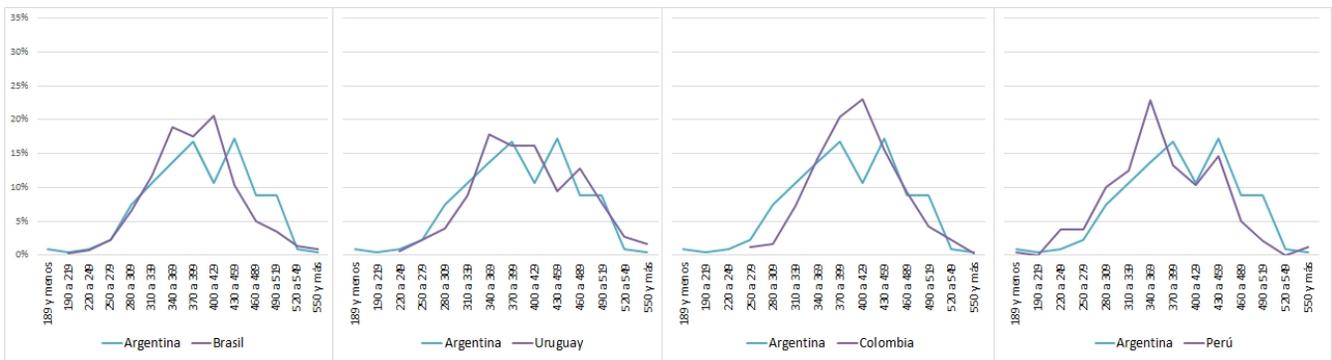
Se resalta particularmente el caso de escuelas de contextos vulnerables que manifestaron resultados muy bajos en 2012. Estas escuelas muestran ciertas características comunes: hay una muy alta presencia de evaluaciones incompletas, bajos porcentajes de respuestas correctas (casi explicadas por el azar), donde el nivel socioeconómico no explica diferencias en su interior. Si bien la información disponible no es suficiente para establecer afirmaciones contundentes, estas características son compatibles con los resultados de escuelas donde la aplicación de la evaluación

despierta escaso interés en estudiantes y docentes. En 2015, las escuelas con similar perfil de vulnerabilidad obtuvieron mejores resultados.

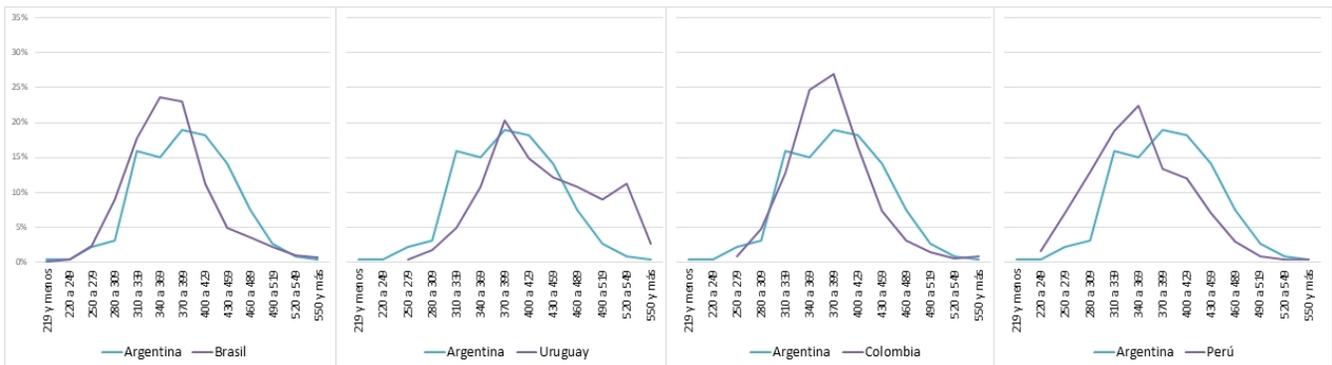
Un último eje de análisis se propone identificar la presencia de grupos similares de “escuelas con muy bajos resultados” en otros países de la región. Se parte del siguiente supuesto: si estos resultados muy bajos se deben principalmente a aprendizajes insuficientes, es esperable que en la distribución de resultados por escuela en otros países también aparezcan estos grupos de escuelas, ya que los niveles de desigualdad son similares (e incluso superiores) a los de Argentina. En cambio, si la existencia de este grupo está asociada a una aplicación “descuidada” de la evaluación (es decir, numerosas pruebas incompletas y tasas de respuestas correctas similares a la distribución por azar), se esperaría que esto no se manifieste en otros países evaluados, al menos en forma generalizada, sino más bien dependiente del contexto. Para la comparación se seleccionan cuatro países que en 2012 tienen similares resultados respecto de Argentina: Brasil, Colombia, Perú, Uruguay.

Gráfico 102. Porcentaje de escuelas según rango de resultados. PISA 2012. Argentina y países seleccionados

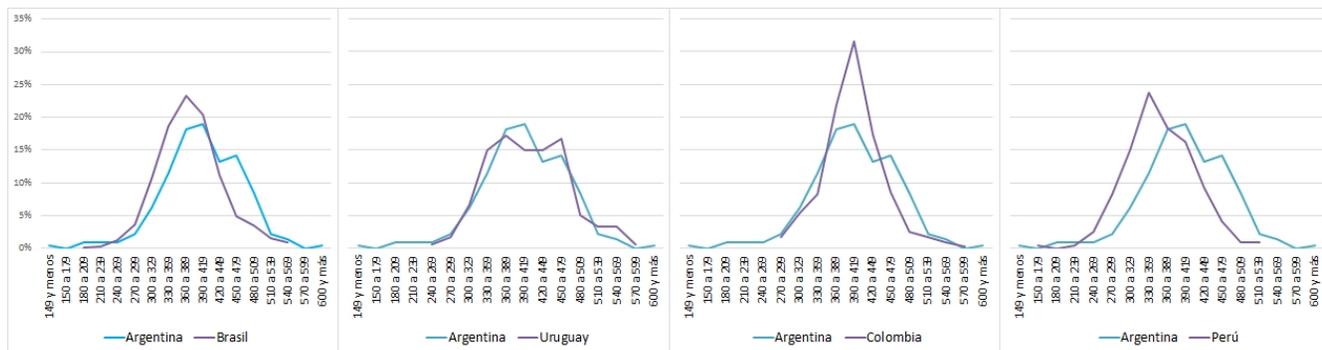
Lectura



Matemática



Ciencias



Fuente: CIPPEC, sobre la base de OCDE PISA 2012 y 2015

Los gráficos presentan la distribución comparada de resultados en escuelas, agrupadas por rangos. Cada gráfico muestra a Argentina comparada con otro país. Se mencionan a continuación los datos más significativos:

- La distribución de los resultados de Brasil, Colombia y Perú tienden a ser más homogéneas que en Argentina, es decir, con mayor concentración de valores medios y menos en los extremos. Esto se reconoce en el hecho de que las curvas de los países tienden a ser más altas (leptocúrticas) que las de Argentina, y los extremos superior e inferior tienden a ser más cortos. En Uruguay, en cambio, no se observa esta particularidad.
- En Uruguay y Colombia no se observa la presencia de “escuelas con resultados muy bajos”. En general, la distribución de escuelas tiende a iniciarse en dos o tres rangos superiores al inicio en Argentina. En el caso de Ciencias, la distancia es mayor entre las escuelas con resultados más bajos de Argentina y de estos países.
- En Brasil y en Perú sí se identifican escuelas con estos resultados (aunque en ambos casos en algunas áreas los resultados mínimos en Argentina son más bajos). Cabe destacar que en estos dos países las curvas están “corridas hacia la izquierda” en relación con Argentina. Esto implica que las escuelas con resultados bajos no se sitúan en valores tan extremos como en el caso de Argentina. En otras palabras, la distancia entre escuelas con resultados bajos y el valor promedio es mucho mayor en Argentina que en Brasil o Perú.
- En síntesis, no se observa en otros países un comportamiento – al menos en forma sistemática – como el que se reconoce en Argentina, de existencia de escuelas con resultados muy por debajo del promedio.

Conclusiones

Los resultados de Argentina para PISA 2015 no se consideran comparables con años anteriores, y esta significativa pérdida de información clave es irremediable. El argumento técnico que fundamenta esta decisión es sólido, y no exige la búsqueda de mayores explicaciones. Esta decisión es consecuencia de omisiones en el marco muestral, a priori no es posible asignarlas a errores técnicos o a exclusiones intencionales. Lo que resulta claramente insuficiente es la explicación de las causas por las que se incurrió en estos errores, y por qué fueron detectados en forma tardía luego de aplicar la prueba y no antes.

Esta decisión también invita a preguntarse por qué los resultados de otros países que muestran también severos problemas de cobertura en sus muestras fueron publicados sin ningún recaudo técnico (tal como se desarrolla en el anexo IV).

Argentina obtuvo en 2015 resultados muy destacados en comparación con ediciones anteriores. Como ambos fenómenos se presentan en forma conjunta, inducen a una rápida (y errónea) asociación entre ambos. Es necesario mirar con cautela los resultados para no llegar a conclusiones apresuradas y erróneas, y explorar posibles explicaciones que den cuenta de esa mejora.

¿Por qué es importante analizar los resultados, si han sido invalidados? Principalmente por las siguientes razones:

- El error en el procedimiento de muestreo no tiene necesariamente relación con los resultados: de no existir sesgos marcados en el grupo de evaluados, es muy posible que los resultados de Argentina en una hipotética muestra correcta alcancen resultados similares a los observados.
- Aunque los resultados no sean comparables estrictamente, no deja de sorprender el nivel de desempeños obtenidos respecto del 2012, crecimiento poco visto en PISA. Si se excluyen explicaciones vinculadas a intervenciones externas que introduzcan sesgos en los datos, es necesario analizar estos resultados para entender la situación reciente del país.

No es posible saber cuáles son las causas de la mejora de los desempeños en la prueba de PISA 2015. Es posible que no haya un solo factor explicativo, pero seguramente algunos de ellos puedan tener mayor peso.

Frente a la imposibilidad de alcanzar certezas, queda la opción de proponer hipótesis, y hacerlas dialogar con la información analizada para sopesar qué evidencias abonan a una u otra opción. Se proponen a continuación tres probables explicaciones, seleccionadas como las más probables dentro del amplio espectro de posibles argumentos que fueron evaluados.

Hipótesis 1 - Muestra manipulada: Esta hipótesis se basa en la afirmación de que existieron intervenciones externas intencionales para incrementar artificialmente los resultados, y que los resultados “reales” de Argentina deberían ser más bajos. Esta hipótesis implica establecer una relación efectiva entre la exclusión de escuelas del marco muestral y los altos resultados: se omitieron escuelas de bajos desempeños para que los resultados sean mejores. Esta argumentación arrojaría un manto de sospecha sobre los resultados de la Ciudad de Buenos Aires, que no fueron invalidados en la publicación. Bajo esta hipótesis, es esperable que los resultados de PISA 2018 se parezcan a 2012, y la aplicación de 2015 quede sólo como ejemplo negativo de manipulación política de resultados. Aunque no alcanzan a descartar esta hipótesis, algunas evidencias ponen en duda sus supuestos:

- El análisis de la muestra de PISA 2015 no evidencia una omisión intencional de escuelas de bajos desempeños. El grupo de estudiantes que la compone es en general similar a 2012, con algunas variaciones menores. E incluso, cuando varían, implican la presencia de mayor población vulnerable en 2015, lo que debiera incidir negativamente en los resultados. Si bien esta equivalencia sólo se realiza sobre un conjunto pequeño de variables (siempre es posible pensar que alguna variable importante queda fuera del modelo), lo cierto es que una planificación a escala de exclusiones intencionales tomaría estos factores en cuenta. Sorprende especialmente el alza en la proporción de escuelas públicas que participaron de la muestra 2015 versus la muestra 2012.
- La distribución de los resultados entre escuelas asume una forma de curva normal, y se comporta de manera bastante similar a 2012 en algunos aspectos clave, como por ejemplo en la relación entre

desempeños y nivel socioeconómico de la escuela. Otras formas de incidencia externa (como ayudas a los estudiantes para identificar las respuestas correctas) dejarían “huellas” en forma de comportamientos atípicos en los datos: concentración de escuelas con desempeños muy destacados, debilitamiento de la relación entre resultados y contexto, distribuciones muy sesgadas hacia altos resultados.

Hipótesis 2 – Mejora educativa: En esta hipótesis no se ponen en duda los resultados de Argentina en PISA 2015. Se afirma que el “salto” evidenciado respecto de PISA 2012 es consecuencia de una mejora en la calidad de los aprendizajes. Los problemas de exclusiones en el marco muestral no afectan sustantivamente a los resultados. Las políticas implementadas en el país en los últimos años habrían ampliado las oportunidades de aprendizaje, y efectivamente PISA muestra que los estudiantes han logrado aprender mejor que en años anteriores. Esta mejora se concentra en el período 2012 - 2015 por sobre años anteriores por razones vinculadas al momento de implementación de las políticas y al tiempo esperable en que manifiesten sus efectos. Bajo esta hipótesis, es esperable que los resultados de PISA 2018 se parezcan a los de 2015, como expresión de una tendencia positiva que permanece instalada en el sistema educativo.

Algunos datos analizados abonan esta interpretación, y otros introducen interrogantes:

- La distribución de los resultados entre escuelas muestra comportamientos esperados a aplicaciones sin manipulación, lo que obligaría a pensar que los resultados son reflejo efectivo de los niveles de aprendizaje de los estudiantes incluidos en las muestras.
- Las muestras no presentan sesgos marcados, lo que invitaría a establecer - sin considerar los requerimientos metodológicos mínimos - cierto marco general de comparabilidad entre años. Esto permitiría poner el énfasis en la mejora efectiva de los desempeños.
- Sin embargo, resulta al menos temerario arriesgar esta hipótesis sin considerar el carácter atípico del incremento, que se ha dado en forma intensa en las tres áreas en simultáneo, y no tiene antecedentes desde la implementación inicial de PISA en el conjunto de países participantes.
- En este sentido, una hipótesis de estas características debiera al menos buscar sustento en un cuidadoso y profundo análisis de las políticas educativas implementadas en la Argentina, y una justificación pedagógica y sociológica que justifique su carácter excepcional, y que adicionalmente puedan ser vinculadas con : (i) una mejora generalizada de los aprendizajes, (ii) en un conjunto de población muy específico (estudiantes de 15 años), (iii) en un momento de tiempo muy específico (exclusivamente a partir del año 2013), y (iv) en todo el país.

Hipótesis 3 – Efecto de la sensibilización a las escuelas: En esta hipótesis, no se ponen en duda los resultados de Argentina en PISA 2015. Los problemas de exclusiones en el marco muestral no afectan sustantivamente a los resultados. Se afirma que el incremento sustantivo de los puntajes respecto de PISA 2012 es consecuencia de una mejora en la calidad de la aplicación de PISA. La principal explicación reside en que en las aplicaciones anteriores no se generaron las condiciones para que los estudiantes respondan a las evaluaciones con su mayor esfuerzo. La falta de incentivos o estímulos para que las escuelas y los estudiantes se comprometan con la evaluación devino en pruebas incompletas y resueltas con desinterés. En esta hipótesis, las acciones de sensibilización desarrolladas en 2015 explicarían la mejora, pero no en términos de “manipulación”, sino en la obtención de información de mayor calidad, que refleja mejor los aprendizajes efectivos. Siendo que estas acciones de sensibilización no son distorsivas, sino por el contrario contribuyen a obtener

resultados más precisos⁵⁸, aquí el “problema” no estaría en 2015, sino en las aplicaciones anteriores que fueron desarrolladas en forma poco cuidada. Bajo esta hipótesis, los resultados de PISA 2018 serían similares a los de 2015 si se garantizan las condiciones mínimas para que los estudiantes respondan con ganas a la prueba. Y los resultados del 2012 serían de baja calidad técnica, ya que habría escuelas donde las pruebas habrían sido muy mal aplicadas.

En general, la información analizada es consistente con esta interpretación:

- La distribución de los resultados entre escuelas muestra comportamientos esperados a aplicaciones sin manipulación.
- Las muestras podrían ser comparables en un sentido general – sin considerar los requerimientos metodológicos mínimos.
- No se observa una mejora en las escuelas con desempeños destacados. El incremento se explica principalmente por (i) el decrecimiento del porcentaje de escuelas con resultados bajos, y por (ii) en 2015 no se observan escuelas con resultados muy bajos.
- El hecho de que en 2015, a diferencia del 2012, no se identifiquen escuelas con muchas pruebas incompletas y con tasas de respuestas correctas similares a una distribución por azar, indicaría una disposición diferente a responder las pruebas en su totalidad: las escuelas vulnerables participantes en 2015 habrían respondido con mayor dedicación.
- No es frecuente que haya escuelas con resultados tan bajos como las identificadas en 2012, con tantas evaluaciones incompletas: en otros países de América Latina con resultados similares en PISA 2012 (Colombia, Brasil, Perú, Uruguay) no se identificó esta distribución. En PISA 2009 ocurre algo similar: aunque los datos no se presentan aquí por motivos de extensión, también allí se identifican escuelas con muy bajos resultados en Argentina, pero no se detectan en otros países con resultados similares.

No es posible ser concluyente en los análisis, y posiblemente nunca se alcance certeza sobre los motivos por los que Argentina muestra ese “salto”. La aplicación de PISA 2018 arrojará mayor luz a estos interrogantes, siempre y cuando el país garantice las condiciones para obtener información de calidad. Hasta entonces, queda la posibilidad de especular sobre estos supuestos y realizar afirmaciones preliminares y tentativas.

Un primer aspecto significativo es que las hipótesis 2 y 3 separan la explicación de los resultados respecto de los problemas en el marco muestral. El análisis desarrollado en el capítulo es consistente con esta apreciación de que se trata de dos fenómenos que podrían ser independientes. Cabe afirmar al respecto que es muy probable que Argentina haya obtenido resultados similares aún sin los errores de omisión del marco muestral.

Un segundo aspecto a considerar en estas hipótesis 2 y 3 es que afirman, bajo diferentes argumentos, que los resultados de Argentina alcanzados en 2015 son un buen reflejo del nivel de aprendizaje de los estudiantes. Bajo estos supuestos, se podría afirmar que los resultados

⁵⁸ Con el término “sensibilización” se hace referencia a las acciones previas a la aplicación de PISA, que se pueden desarrollar en las escuelas, y que contribuyen a promover que las evaluaciones sean respondidas por los estudiantes con el mayor esfuerzo, concientizando sobre su importancia y su funcionamiento. En el capítulo 3 del documento se aborda el tema de la sensibilización con mayor profundidad.

posicionarían al país en el grupo de desempeños más destacados, cerca de Uruguay y México, y por debajo de Chile.

Un tercer punto pone en el centro de la escena las acciones de “sensibilización” a las escuelas, sobre el cual es necesario arrojar luz. No se debe confundir la implementación de acciones en las escuelas que favorezcan una mayor calidad de las respuestas, respecto de aquellas que distorsionan los resultados. La expresión “preparar a las escuelas” carga con esa ambigüedad de interpretaciones.

Es posible establecer criterios técnicos rigurosos para distinguir una situación de otra. El trabajo de sensibilización se enfoca en las escuelas y estudiantes para que conozcan y dimensionen la importancia de la evaluación, para que se familiaricen con los formatos de la prueba y se promueva que respondan la mayor cantidad posible de ítems. De esta forma, se promueve que las pruebas se respondan con esfuerzo y reflejen lo mejor posible los saberes efectivos de los estudiantes.

La aplicación de PISA 2018 deberá contemplar este desafío, así como también tendrá que considerar la implementación de criterios de transparencia que permitan responder muchos de los interrogantes que la aplicación de PISA 2015 ha dejado abiertos.

Anexo III - Incidencia de los ítems NRI en el porcentaje de respuestas correctas de PISA 2006 - 2015

Como fue presentado en el capítulo 2, el consorcio PISA implementó en 2015 cambios en la metodología de escalamiento de resultados. Entre otros aspectos, modificó el modo en que se consideraban los *non-reached items* (NRI) para el cálculo de los resultados. Por lo tanto, es esperable que los países donde la presencia de estos ítems en los resultados es más alta se vean más afectados por el cambio de criterio.

Para poner en evidencia este efecto, el consorcio PISA elaboró un ejercicio de cálculo: tomando el conjunto de ítems utilizados en dos aplicaciones diferentes, contabilizó el porcentaje de respuestas correctas si se consideran a los ítems NRI como respuestas incorrectas. Luego, realiza el mismo cálculo pero considerando los ítems NRI como no administrados. De esta forma, representa en qué medida los cambios en la interpretación de los ítems NRI pueden afectar a los resultados globales.

El siguiente cuadro presenta estos datos, indicando para cada país el porcentaje de ítems respondidos en forma correcta según la metodología implementada en 2006 - 2012 (es decir, considerando a los NRI como incorrectos), que se presentan en el cuadro bajo el título de "Método A". También se incluye el "Método B", que muestra el porcentaje de ítems correctos según la metodología de PISA 2015 (es decir, considerando a los NRI como no administrados).

La comparación entre ambos valores permite dimensionar el impacto de este cambio de metodología en los resultados de PISA: el porcentaje de respuestas correctas a la evaluación tiene una relación directa con el puntaje promedio obtenido por el país en cada área.

Cuadro 5. Porcentaje de ítems con respuestas correctas, calculadas según método A y B, por área y país. Años 2006-2015

		Lectura			Matemática			Ciencias				
		Mét. A	Mét. B	Dif.	Mét. A	Mét. B	Dif.	Mét. A	Mét. B	Dif.		
ARG ^(*)	2009	43,6	49,2	5,6	2012	26,4	27,7	1,3	2006	30,8	34,8	4,0
	2015	60,3	63,3	3,0	2015	39,0	41,7	2,7	2015	45,1	47,6	2,5
	2015-2009	16,7	14,1	-2,6	2015-2012	12,6	14,0	1,4	2015-2006	14,3	12,8	-1,5
BRA ^(**)	2009	45,3	47,6	2,3	2012	26,7	28,2	1,5	2006	30,2	32,4	2,2
	2015	47,3	48,1	0,8	2015	25,2	25,6	0,4	2015	32,2	32,8	0,6
	2015-2009	2,0	0,5	-1,5	2015-2012	-1,5	-2,6	-1,1	2015-2006	2,0	0,4	-1,6
CHI	2009	54,5	56,8	2,3	2012	31,5	32,4	0,9	2006	38,8	41,3	2,5
	2015	58,6	59,3	0,7	2015	32,0	32,4	0,4	2015	39,6	40,4	0,8
	2015-2009	4,1	2,5	-1,6	2015-2012	0,5	-	-0,5	2015-2006	0,8	-0,9	-1,7
COL	2009	45,6	51,3	5,7	2012	24,0	26,6	2,6	2006	30,7	35,9	5,2
	2015	50,3	51,1	0,8	2015	26,2	26,7	0,5	2015	33,8	34,5	0,7
	2015-2009	4,7	-0,2	-4,9	2015-2012	2,2	0,1	-2,1	2015-2006	3,1	-1,4	-4,5
CRI	2009	51,5	54,5	3,0	2012	34,1	36,3	2,2				
	2015	50,8	51,1	0,3	2015	33,7	34,4	0,7				
	2015-2009	-0,7	-3,4	-2,7	2015-2012	-0,4	-1,9	-1,5				
MEX	2009	48,7	52,0	3,3	2012	30,7	32,8	2,1	2006	33,7	35,9	2,2
	2015	50,8	51,1	0,3	2015	30,1	30,3	0,2	2015	33,9	34,2	0,3
	2015-2009	2,1	-0,9	-3,0	2015-2012	-0,6	-2,5	1,9	2015-2006	0,2	-1,7	-1,9
PER	2009	37,5	42,3	4,8	2012	23,3	25,9	2,6				
	2015	45,8	46,1	0,3	2015	27,0	27,1	0,1				
	2015-2009	8,3	3,8	-4,5	2015-2012	3,7	1,2	-2,5				
URU	2009	49,4	53,9	4,5	2012	30,1	32,8	2,7	2006	37,7	41,2	3,5
	2015	52,1	53,2	1,1	2015	30,9	31,5	0,6	2015	37,5	38,3	0,8
	2015-2009	2,7	-0,7	-3,4	2015-2012	0,8	-1,3	-2,1	2015-2006	-0,2	-2,9	-2,7

(*) Los datos de Argentina fueron incluidos a modo ilustrativo, aunque no se consideran comparables debido a que la muestra no cubrió la población total objetivo, a causa de una posible omisión de las escuelas del marco muestral.

Nota: estimación realizada sobre aproximadamente 58 ítems de anclaje de ciencias, 50 de lectura y 41 de matemática. El total de ítems considerados puede variar entre países

(**) Los datos de Brasil en el 2012 no representan a toda la población, ya que se excluyeron a las escuelas rurales plurigrado.⁵⁹

Fuente: CIPPEC sobre la base de OCDE (2017) Trends in the percentage of correct answers and treatment of non-reached items (Mimeo).

⁵⁹ Para mayor referencia, puede consultarse el anexo A7 de OCDE (2014) PISA 2012 Results: What Students Know and Can Do. Student Performance in Mathematics, Reading and Science (Volume I)

Tal como puede observarse, los resultados del año 2015 son similares en ambos métodos: a excepción del caso de Argentina, y en lectura para el caso de Uruguay, en el resto de los países el porcentaje varía menos de un punto porcentual. Esto se debe a que en el año 2015, en la mayoría de los países de la región, el porcentaje de ítems NRI es muy bajo. Por lo tanto, no se observa una diferencia significativa al considerarlos como erróneos o como no administrados.

En cambio, puede observarse que en los años anteriores, el porcentaje de respuestas correctas varía mucho en los países, según la forma en que se consideran los ítems NRI. El análisis de algunos ejemplos puede contribuir a ilustrar mejor estas diferencias.

El caso del área de lectura en Colombia es paradigmático: si los ítems NRI se consideran erróneos (Método A), la comparación en el tiempo pone en evidencia que el país mejora en sus resultados, incrementando el % de respuestas correctas del 45,6% al 50,3% (considerando el universo de ítems de anclaje). Sin embargo, si los ítems NRI se consideran como no administrados, esta mejora desaparece: el porcentaje de ítems respondidos en forma correcta sería del 51,3% en 2009 y del 51,1% en 2015, por lo tanto, no sería posible afirmar que en ese periodo el país mejoró sus desempeños. Es decir, la mejora que puede reconocerse en el país en esta área, al menos para los ítems de anclaje, se explica en su totalidad por la reducción de la cantidad de ítems que quedan en blanco al final de la evaluación. Si se excluyen estos casos, los estudiantes respondieron en forma similar ambas evaluaciones.

Otro caso paradigmático para analizar como ejemplo es el de Uruguay 2006–2015 en el área de ciencias. Aquí, considerando el universo de 58 ítems de anclaje, el país muestra porcentajes de respuestas correctas similares en ambas evaluaciones considerando el método A (37,7% en 2006 y 37,5% en 2015). Es decir, si los ítems NRI se consideran incorrectos, en ambas evaluaciones los resultados son similares. Ahora bien, si los ítems NRI se excluyen del análisis (considerados no administrados por el método B), lo que se observa en el país es una disminución del porcentaje de respuestas correctas. Es decir, para este conjunto de ítems analizados, los estudiantes uruguayos cometieron más errores en 2015 respecto de 2006, pero si se consideran a los ítems NRI como erróneos estas equivocaciones se compensan y desaparecen las diferencias.

No hay una relación simple que permita calcular cuánto impacta en el resultado de PISA (en puntaje promedio) estas variaciones en el porcentaje de respuestas correctas, ya que depende de las características de los ítems. Los resultados reescalados consideran todos los cambios implementados en la metodología; no es posible aislar el efecto de los ítems NRI sobre otros aspectos. Como referencia general, se puede citar aquí el cálculo de Wuttke (2007), que estima para PISA 2003 una variación promedio de 4 puntos en la escala PISA en una diferencia del 1% en el porcentaje de respuestas correctas.

Anexo IV Interrogantes sobre la calidad de las muestras de PISA en algunos países

El programa PISA se plantea como objetivo principal evaluar la “calidad, equidad y eficiencia de los sistemas escolares de los países participantes” (OCDE, 2006). Esta definición es central para entender algunos aspectos de PISA: la construcción de conocimiento sobre los saberes de los estudiantes no es un fin en sí mismo, sino un medio para enunciar características y propiedades de los sistemas educativos nacionales.

Para lograr este objetivo, implementa un conjunto de pruebas e instrumentos destinados a construir una medición sobre la formación de los alumnos cuando llegan al final de la etapa de enseñanza obligatoria, hacia los 15 años de edad. El foco de esta medición es producir información orientada a la toma de decisiones de política educativa de los países (OECD, 2006).

Para alcanzar los niveles de calidad necesarios para poder realizar estas afirmaciones sobre la calidad, equidad y eficiencia de los sistemas educativos, son tres los requisitos centrales que debe cumplir la información generada por PISA:

- **Generalización:** Las conclusiones obtenidas sobre los resultados de los estudiantes evaluados en PISA deben ser válidas para toda la población objetivo, para ello deben cumplirse una serie de requisitos. Cualquier factor que afecte a la representatividad de la muestra pone en duda que sea posible trasladar sus resultados a la población del país.
- **Coherencia:** Son aquellos aspectos de la implementación de la evaluación que garantizan su comparabilidad, en tanto permitan obtener datos equivalentes en todos los países.
- **Precisión:** Refiere al tratamiento de los errores en pos de generar información de calidad. Se trata de construir estimaciones con bajos niveles de error en la medición, pero también reducir la posible incidencia de errores no estadísticos o aleatorios, que pueden distorsionar los resultados.
- (OCDE, 2015)

Las propiedades de precisión, coherencia y generalización de los resultados son aspectos centrales del programa PISA. Todo el edificio analítico de PISA se construye en base a la afirmación de que los resultados son representativos de los países, y comparables entre sí con altos niveles de precisión.

Para garantizar que la información que se produce en cada operativo cumpla estas tres propiedades, las acciones de PISA se planifican y desarrollan en base a un conjunto de estándares técnicos formulados por la OCDE. Estos estándares conforman criterios ordenadores que estructuran las acciones de los técnicos involucrados en las diferentes etapas de cada evaluación. El cumplimiento de estos estándares constituye un compromiso mutuo de las partes, y existen diferentes instancias de monitoreo y seguimiento de su concreción efectiva.

PISA establece tres tipos de estándares, cada uno con un propósito específico: estándares de datos, que se orientan a garantizar la calidad de la información recolectada; estándares de gestión, para monitorear que todos los objetivos operativos de PISA se cumplan de manera oportuna y coordinada; y estándares nacionales de calidad, vinculados a los procesos que acreditan la equivalencia cultural y lingüística del contenido de las evaluaciones.

Por lo tanto, la calidad de la información producida por PISA es consecuencia de un compromiso articulado entre los responsables internacionales del diseño y ejecución de cada evaluación (tanto a nivel de la OCDE como de las empresas contratadas como consultores

internacionales), y el equipo técnico nacional de cada país que asume la responsabilidad de ejecutar, supervisar y monitorear las acciones previstas para cada operativo⁶⁰.

Un aspecto central a considerar en materia de esta serie de procesos y controles que se ponen en juego en la construcción de resultados comparables entre países es la definición de la población objetivo a evaluar. PISA establece grupos de población equivalentes a partir de la edad. Justifica esta elección –en lugar de otros criterios como el grado- considerando que permite construir poblaciones más comparables: “Las diferencias entre países en la naturaleza y duración de la educación infantil, la edad de acceso a la enseñanza obligatoria y la estructura del sistema educativo hacen que los cursos escolares difieran e impidan la comparación a escala internacional” (OCDE 2010).

Es decir, se evalúa a los estudiantes de 15 años independientemente de la trayectoria educativa que hayan tenido, considerando que PISA no evalúa saberes escolares sino competencias para la vida. Esta definición ha despertado diferentes reacciones de oposición, basadas principalmente en el argumento de que se agrupa población con desigual tiempo de permanencia en la educación. Esto ocurre tanto por las diferencias en la edad de ingreso al sistema educativo (que implica distinta cantidad de años de escolarización para la misma edad), como por la acumulación de repitencia (que implica igual cantidad de años de escolarización pero menor exposición al aprendizaje del currículum). Por otro lado, implica una desvinculación del aprendizaje respecto de la enseñanza: los estudiantes de 15 años evaluados asisten a diferentes grados, por lo que no es posible construir información que permita vincular los resultados a la tarea de los docentes, a partir de un instrumento destinado a ellos (Carnoy, 2015).

En la negociación que establece el consorcio PISA con cada uno de los países participantes, esta definición de población objetivo se divide en dos conceptos clave:

- La *población objetivo deseada* es aquella que se propone evaluar PISA: los estudiantes de entre 15 años y 3 meses a 16 años y 2 meses de edad, (considerando como fecha de referencia el inicio del período de exámenes), que asiste a 7mo grado de estudios o más, en instituciones educativas ubicadas dentro del país.
- La *población objetivo definida* es el conjunto de los estudiantes que, al momento de seleccionar la muestra, poseen una probabilidad estadística de ser evaluados. Es decir, todos los que conforman el marco muestral de PISA.

La diferencia entre la población objetivo deseada y definida está dada por las exclusiones: estudiantes que no se incluyen en el marco muestral bajo determinados argumentos que cada país puede negociar con el consorcio PISA. Estas exclusiones pueden justificarse en función de ciertas características de la población (por ejemplo, estudiantes con ciertos tipos de discapacidad) o de las escuelas (por ejemplo, establecimientos muy pequeños, o de muy difícil acceso). Pero no pueden superar el 5% de la población objetivo definida (Estándar Técnico de PISA Nro. 1.7).

La población objetivo definida es la que conforma el marco muestral de PISA en cada país. De allí se eligen las escuelas que serán evaluadas. La muestra construida será representativa de este

60 Este último punto es fundamental. Existe un requisito implícito en la adopción del programa PISA, que es la necesidad de que los países hayan alcanzado un nivel alto de desarrollo de su capacidad institucional para poder implementar los sucesivos pasos que requiere la evaluación de forma tal que se garantice la calidad de la información (Solano-Flores y Milbourn, 2016). La calidad de los resultados de PISA depende, entre otros aspectos, de que los técnicos de cada país asignados a diferentes responsabilidades (ya sea en la elaboración del marco muestral, en la aplicación en territorio, en la corrección de ítems abiertos, entre otros pasos clave) puedan desarrollar sus tareas en forma correcta, manteniendo los estándares de calidad previstos, y en el tiempo planificado.

universo conformado por el marco muestral. En este proceso, los equipos nacionales a cargo de la implementación son los responsables de construir el marco muestral, y de definir las exclusiones en negociaciones con el consorcio PISA. PISA se reserva la selección de escuelas y estudiantes a evaluar, a partir de este marco.

El consorcio PISA realiza diversos controles sobre la cobertura de la muestra. Existe un conjunto amplio de definiciones que establecen las condiciones que la muestra debe cumplir para considerarse válida. Estos criterios se expresan en sus estándares técnicos. Entre ellos, es relevante hacer referencia a algunos en particular.

Estándar 1.1 La Población Objetivo Deseada de PISA se acuerda mediante negociación entre el Gerente de Proyecto Nacional y el Contratista Internacional, dentro de las restricciones impuestas por la definición de la Población Objetivo de PISA.

Estándar 1.4 Las escuelas son muestreadas utilizando principios acordados de muestreo científico, profesionalmente establecidos y reconocidos.

Estándar 1.6 Los estudiantes son muestreados usando principios acordados, establecidos y reconocidos profesionalmente de muestreo científico y de una manera que representa a toda la población de estudiantes elegibles para PISA.

Estándar 1.7 La Población Objetivo Definida de PISA cubre el 95% o más de la población objetivo deseada PISA. Es decir, las exclusiones en el nivel escolar y las exclusiones dentro de la escuela combinadas no exceden el 5%.

Estándar 1.10 La tasa de participación final ponderada de las escuelas debe ser al menos el 85% de la muestra. Si la tasa de respuesta es inferior al 85%, se puede alcanzar una tasa de respuesta aceptable mediante el uso acordado de las escuelas de reemplazo.

Estándar 1.11 La tasa de respuesta final ponderada de los estudiantes debe ser al menos el 80% de todos los estudiantes de la muestra.

Estándar 1.13 A menos que se acuerde lo contrario, el Contratista Internacional responsable seleccionará la muestra de escuelas para la evaluación.

El estándar 1.7 resguarda la calidad del marco muestral, garantizando que el mismo (que contiene a la Población Objetivo Definida) cubra al menos al 95% de la Población Objetivo Deseada. Los estándares 1.4 y 1.13 resguardan la objetividad del proceso de selección de la muestra, realizado por entidades internacionales y bajo criterios científicos de calidad. Los estándares 1.7, 1.10 y 1.11 se formulan para controlar la participación efectiva de las escuelas (y de los alumnos evaluados dentro de ellas), de modo tal que no se presenten sesgos ocasionados por escuelas o alumnos no evaluados.

A primera vista, este conjunto de criterios combinados parece proteger las principales aristas de los factores de riesgo que pueden afectar la representatividad de la muestra, al menos en lo que respecta a la cobertura de la población objetivo.

Para el control de calidad de este proceso, PISA presenta en el informe 2015 tres índices de cobertura:

Índice de cobertura 1: Es la diferencia entre la población objetivo deseada y definida. Es decir, excluyendo a aquellas poblaciones que el país opta por no evaluar. Valores bajos en este indicador señalan que el país decidió excluir poblaciones o minorías en base a criterios generales que dejan fuera a muchos estudiantes de 15 años. Este valor no puede ser inferior al 95%.

Índice de cobertura 2: Es la diferencia entre la muestra seleccionada efectiva sobre la población esperada de PISA. En otras palabras, a la medición anterior se le agrega la consideración de las exclusiones a nivel de alumno al momento del campo. Valores bajos en este indicador señalan que el país se encontró con muchos estudiantes que no pudo evaluar al momento de aplicar el campo (por

inasistencia, abandono, por tratarse de estudiantes con discapacidad, u otras razones). Este valor tampoco puede ser inferior al 95%.

Índice de cobertura 3: Cobertura respecto de la población: Se relaciona la matrícula ponderada con la población de 15 años, para medir el efecto de la población que no asiste. Valores bajos en este indicador pueden señalar tanto problemas de cobertura de la muestra, o países con mucha población de 15 años fuera de la escuela.

Los dos primeros índices tienen vinculación con los estándares de calidad formulados por PISA. El tercer indicador no da cuenta por sí sólo de un problema de calidad de la muestra, sino que evalúa la representatividad de los resultados de PISA respecto de la población total del país.

Cuadro 14. Índices de cobertura 1,2 y 3 de PISA 2015. Países de América Latina

	Índice de cobertura 1	Índice de cobertura 2	Índice de cobertura 3
Argentina	0,992	0,992	0,550
Brasil	0,972	0,972	0,707
Chile	0,983	0,982	0,798
Colombia	0,999	0,999	0,746
México	0,991	0,991	0,617
Perú	0,985	0,985	0,744
Uruguay	0,999	0,999	0,715

Fuente: CIPPEC sobre la base de OCDE, PISA 2015

Las dos primeras columnas muestran que, a primera vista, los datos de cobertura de PISA 2015 de los países de América Latina cumplen con los estándares técnicos requeridos, ya que el total de exclusiones en ningún país alcanza al 5% de la población objetivo. A simple vista parece que todos los países cumplen con el estándar de calidad de cobertura de PISA.

El tercer indicador de cobertura relaciona la muestra con la población total de 15 años, para evaluar en qué medida los resultados son representativos del conjunto. Este indicador debe interpretarse con mucho cuidado, ya que no se trata de un indicador de calidad de la muestra, sino de representatividad de los resultados respecto de la población total.

El valor del indicador está mediado por los niveles de cobertura del sistema educativo. No distingue entre países que puedan tener problemas de cobertura de PISA respecto de aquellos con altos niveles de exclusión educativa. Un ejemplo puede ilustrar mejor este problema, tomando la información de cobertura publicada por PISA.

Cuadro 15. Población de 15 años, población objetivo de PISA e índice de cobertura 3. PISA 2015. Países de América Latina

	[1]	[2]	[3]	[4]	[5]
	Población de 15 años	Población objetivo deseada <i>Población de 15 años en el grado 7 o más</i>	Población objetivo deseada/ población total	Índice de cobertura 3	Población objetivo - índice de cobertura 3 (3) - (4)
Argentina	718.635	578.308	0,805	0,550	0,255
Brasil	3.430.255	2.853.388	0,832	0,707	0,125
Chile	255.440	245.947	0,963	0,798	0,165
Colombia	760.919	674.079	0,886	0,746	0,140
México	2.257.399	1.401.247	0,621	0,617	0,004
Perú	580.371	478.229	0,824	0,744	0,080
Uruguay	53.533	43.865	0,819	0,715	0,104

Fuente: CIPPEC sobre la base de OCDE, PISA 2015

El cuadro presenta el porcentaje de la población de 15 años de un país que compone la población objetivo deseada de PISA (columna 3). Cabe recordar que, para definir a la población objetivo, se consideran dos criterios clave: que asista al sistema educativo, y que se encuentre inscripto al menos en el grado 7.

El índice de cobertura 3 relaciona la muestra ponderada con la población total, como fue explicado previamente. Esta relación está necesariamente mediada por la cobertura del sistema educativo: un índice bajo de cobertura podría indicar tanto a un país con una muestra poco representativa de la población objetivo, o un país con una buena muestra pero con bajos niveles de cobertura del sistema educativo (o con muchos estudiantes de 15 años asistiendo a grados inferiores al 7^{mo})

Por ejemplo, en México el 61,7% de la población de 15 años está representada en la muestra (columna 4). ¿Esto indica un problema en la muestra? No es posible afirmarlo ni negarlo. Es necesario evaluar el tamaño de la población objetivo. En México el porcentaje de adolescentes de 15 años fuera de la escuela es muy alto, ellos no participan de PISA porque no conforman su población objetivo. En la columna 3 puede observarse que solamente el 62,1% de la población constituye objetivo de la evaluación PISA⁶¹. Por lo tanto, en este país, un porcentaje bajo del indicador de cobertura 3 no refleja problemas de cobertura o calidad de la muestra, sino perfila un sistema educativo con poca cobertura. Para facilitar la comparación, se inserta la columna 5 -de elaboración propia - con el cálculo de la diferencia entre ambos indicadores analizados.

61 Es importante considerar que los valores presentados y analizados aquí son cifras preliminares. Los datos de población total y población objetivo son estimaciones elaboradas por los países uno o dos años antes de la implementación, ya que requeridos por PISA como insumos para el armado del marco muestral. Es frecuente que el país actualice posteriormente estas cifras, lo que provoque un cambio en los totales. En el caso particular de México, PISA advierte que el valor de población objetivo es una estimación basada en el marco de muestra. México informó posteriormente la cifra oficial, que fue de 1.573.952 (OCDE 2016, pág. 292), por lo que la población objetivo representaría el 69,7% de la población total.

Distinto es el caso de Argentina: el país muestra una baja relación entre la muestra ponderada y la población de 15 años (55,0%). Sin embargo, la población objetivo de PISA engloba al 80,5% de los adolescentes de 15 años en el país. La diferencia entre ambos valores estaría indicando la existencia de algunos problemas.

El informe de PISA 2015 también incluye los valores de cobertura de los países utilizados para el monitoreo del cumplimiento de los estándares de calidad establecidos por el consorcio PISA.

Cuadro 66. Indicadores de cobertura vinculados a los estándares técnicos de calidad de PISA. PISA 2015

	[1]	[2]	[3]	[4]	[5]	[6]
	% de exclusiones a nivel de escuela	% de exclusiones dentro de las escuelas	% global de exclusión[*]	Tasa ponderada de participación de las escuelas [antes de los reemplazos]	Tasa ponderada de participación de las escuelas [luego de los reemplazos]	Tasa ponderada de participación de los estudiantes [luego de los reemplazos]
Argentina	0,45%	0,34%	0,80%	88,74%	97,13%	90,36%
Brasil	2,26%	0,56%	2,80%	93,19%	94,08%	87,32%
Chile	1,07%	0,68%	1,75%	92,43%	99,14%	93,31%
Colombia	0,01%	0,09%	0,09%	98,64%	99,81%	94,52%
México	0,42%	0,49%	0,91%	95,46%	97,52%	95,43%
Perú	1,33%	0,17%	1,50%	99,52%	99,79%	98,90%
Uruguay	0,01%	0,08%	0,09%	98,28%	99,33%	86,16%

[*] El porcentaje global de exclusión se calcula como el porcentaje de exclusiones a nivel de escuela + dentro de la escuela menos 1 por el porcentaje de exclusiones a nivel de escuela, es decir: $[1] + [2] * (1 - [1])$

Fuente: CIPPEC sobre la base de OCDE, PISA 2015

La columna 3 hace referencia directa al estándar de calidad 1.7 de PISA, dado que permite monitorear si las exclusiones de un país superan al 5% de la población objetivo. Según este cuadro, todos los países de la región se han mantenido dentro de los límites previstos. Las columnas 4 y 5 permiten monitorear el cumplimiento del estándar 1.10, sobre la tasa de participación mínima (85%) de las escuelas de la muestra, antes y después de los reemplazos. La columna 5 se relaciona con el estándar 1.11 que establece que la tasa ponderada de participación de los estudiantes de la muestra debe ser, al menos, del 80%.

Tal como es posible observar en el cuadro, todos los países de la región cumplen con los requisitos establecidos en los estándares de calidad, y en general con amplios márgenes respecto de los límites impuestos. Sin embargo, el caso de Argentina 2015 resulta ser una llamativa alerta sobre el alcance de estos estándares para poder dar cuenta de problemas de cobertura en la muestra. De alguna forma que no se define con claridad en el informe, el problema de la cobertura de la muestra en Argentina no se ve reflejado en estos indicadores de monitoreo de la calidad.

Cabe señalar al respecto que PISA incluye dos indicadores adicionales de cobertura (Índices de cobertura 4 y 5) que podrían arrojar más luz sobre el problema. Estos índices no forman parte de las publicaciones de difusión de los resultados de PISA, pero sí se incluyen en el reporte técnico. El índice de cobertura 4 caracteriza la relación entre (i) la estimación ponderada de la muestra de

estudiantes, más el total ponderado de estudiantes que fueron excluidos en la aplicación en las escuelas, y (ii) el total de estudiantes de 15 años que surge del marco muestral. Este indicador permite evaluar la representatividad de la muestra respecto de la población que se propone caracterizar.

En índice de cobertura 5 muestra la relación entre el total de estudiantes del marco muestral, y el total de población objetivo menos las exclusiones de escuelas. De esta manera, es posible evaluar en qué medida el marco muestral representa a la población que se definió como elegible para PISA.

Cuadro 17. Índices de cobertura 4 y 5 de PISA 2015, y sus componentes. Países de América Latina

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
	Población objetivo deseada	Exclusiones de escuelas	Estudiantes en el marco muestral	Muestra ponderada	Exclusiones al interior de las escuelas	Índice de cobertura 4 ([4] + [5]) / [3]	Índice de cobertura 5 ([1] - [2]) / [3]
Argentina	578.308	2.617	572.941	394.917	1.367	0,69	1,00
Brasil	2.853.388	64.392	2.692.686	2.425.961	13.543	0,91	0,97
Chile	245.852	2.641	232.756	203.782	1.393	0,88	0,96
Colombia	674.079	37	673.817	567.848	507	0,84	1,00
México	1.401.247	5.905	1.373.919	1.392.995	6.810	1,02	0,98
Perú	478.229	6.355	470.651	431.738	745	0,92	1,00
Uruguay	43.865	4	43.737	38.287	32	0,88	1,00

Fuente: CIPPEC sobre la base de datos de OCDE, PISA 2015

El indicador de cobertura 4 se basa en la relación entre la muestra ponderada y la población objetivo definida, aquella que se incluye en el marco muestral. Es decir, establece qué proporción de la población sobre la que se deseaba construir una medición de los aprendizajes estuvo efectivamente representada, incluyendo a las exclusiones que se definieron al momento de la aplicación de las escuelas.

Por lo tanto, esta medida constituye una aproximación que resume varios de los estándares postulados por PISA en un único valor: en qué medida la muestra se aproxima o se aleja de aquella población a la que se desea caracterizar. La existencia de una brecha entre la población objetivo definida (marco muestral) y la muestra ponderada pone en foco la existencia de ciertas dificultades en la implementación de la evaluación.

Cabe recordar que el marco muestral está compuesto por el conjunto de estudiantes que *podrían ser* seleccionados para la evaluación de PISA. La teoría del muestreo establece que la muestra seleccionada debe ser representativa de la población incluida en el marco muestral. Los factores de ponderación asignados a cada caso evaluado para ponderar la muestra se basan (aunque luego se realizan ajustes) en la probabilidad de selección definida en el marco muestral.

Es decir, la existencia de amplias diferencias entre el total de estudiantes del marco muestral y la muestra ponderada está indicando la presencia de algún problema que pone en juego la representatividad de los resultados, es decir, en qué medida son representativos de la población a la que desean representar.

Tal como puede observarse en la columna 6, la muestra de México representa prácticamente a la población objetivo (incluso la supera en un 0,02%), por lo que se trataría de una muestra con muy altos niveles de cobertura. En cambio, los niveles de cobertura de Argentina son muy bajos, apenas del 69%. Es decir, aproximadamente uno de cada tres estudiantes de 15 años no esté representado en la muestra. En el informe de PISA 2015 no se establece una relación explícita entre este indicador de cobertura y la decisión de establecer que los resultados de Argentina no son representativos ni comparables. Sin embargo, este indicador es el que establece una alerta de que podría haber problemas de cobertura de la muestra de PISA.

Las variaciones en el tiempo de este indicador permiten identificar qué aplicaciones de PISA han manifestado mayores dificultades en los países, como se observa en el siguiente cuadro.

Cuadro 18. Índice de cobertura 4 de PISA. Países de América Latina

	año 2000(*)	año 2003	año 2006	año 2009	año 2012	año 2015
Argentina			91%	78%	90%	69%
Brasil	97%	84%	80%	80%	90%	91%
Chile			94%	95%	96%	88%
Colombia			100%	93%	92%	84%
México	90%	90%	89%	93%	92%	102%
Perú				89%	82%	92%
Uruguay		79%	90%	78%	87%	88%
Promedio América Latina		84%	91%	87%	90%	88%
Promedio OCDE	94%	95%	96%	95%	96%	97%

(*) Sólo se dispone de información para PISA 2000 de los países OCDE

Fuente: CIPPEC sobre la base de OCDE, PISA 2015

El **cuadro 18** permite identificar diferencias entre países en relación con la cobertura de la población definida a lo largo del tiempo. En promedio, se ha mantenido estable a lo largo del tiempo, en torno al 85%, unos siete puntos por debajo del promedio OCDE. Sin embargo la situación de los países es muy diferente: casos como México, Brasil, Uruguay y Perú han mejorado a lo largo del tiempo la representatividad de la muestra. Chile y Colombia muestran una caída en la muestra del año 2015, aunque en niveles similares al resto de los países. Argentina muestra un comportamiento oscilante, con niveles muy bajos en los años 2009 y 2015.

En los estándares técnicos de PISA, no se establecen valores ni niveles mínimos de calidad para este indicador. La ausencia de un estándar de calidad específico para esta brecha no permite establecer cuáles son los límites para establecer que un valor del índice 4 pone en duda la representatividad de la muestra.

El caso de Argentina 2015 establece una alerta sobre la necesidad de evaluar con mayor profundidad este índice. En la historia de las aplicaciones de PISA, ha habido otros casos de países donde se identificó una baja cobertura del índice 4, en ciertos años de implementación de PISA. En estos casos, las muestras parecen estar bastante alejadas de la población a la que se proponía representar.

Cuadro 19. Índices de cobertura 1 a 5 de PISA. Países seleccionados

País - Año	Índice de cobertura 1	Índice de cobertura 2	Índice de cobertura 3	Índice de cobertura 4	Índice de cobertura 5
Costa Rica - 2012	100%	100%	50%	62%	101%
Azerbaiyán - 2009	99%	99%	57%	63%	92%
Turquía- 2003	99%	99%	36%	67%	100%
Argentina - 2015	99%	99%	55%	69%	100%
Líbano - 2015	98%	98%	66%	70%	100%
Suiza - 2000	98%	98%	89%	75%	124%
Panamá - 2009	99%	99%	53%	75%	94%
Trinidad y Tobago - 2015	100%	100%	76%	76%	100%
Serbia - 2003	94%	94%	69%	76%	103%
Portugal - 2012	98%	98%	88%	76%	100%
Costa Rica - 2015	100%	100%	63%	78%	101%
Argentina - 2009	99%	99%	69%	78%	96%
Uruguay - 2009	100%	100%	63%	78%	100%
Uruguay - 2003	100%	100%	63%	79%	107%

Fuente CIPPEC sobre la base de OCDE, PISA 2015

Estos países son los que muestran los niveles más bajos del índice de cobertura 4 en las aplicaciones de PISA 2000 - 2015. Se estableció arbitrariamente presentar los casos donde la muestra ponderada no alcanza a representar al 80% de la población incluida en el marco muestral.

En todos los casos, los índices 1 y 2 muestran valores muy altos, lo que no permite identificar problemas vinculados a los estándares técnicos de calidad formulados por PISA. El índice 5 tampoco presenta para ninguno de los países valores preocupantes. Respecto del índice 3, como ya fue desarrollado en forma previa, por sí sólo no indica problemas o dificultades en la cobertura de PISA.

Los países representados en esta lista han evaluado, en los años indicados, a una muestra de estudiantes que - ponderada - deja fuera a, al menos, el 20% de los estudiantes que conformaron la población objetivo definida del país.

No es posible establecer con claridad cuáles son los motivos por los que se manifiestan estas brechas: no se ha identificado en los informes de PISA información o aclaraciones metodológicas al respecto. Estos bajos niveles de cobertura no son mencionados ni referidos como problemáticos, ni se realizan advertencias sobre sus implicancias para la interpretación de los resultados.

En un muestreo sin errores, la muestra ponderada siempre se asemeja a la población del marco muestral. Está dentro de los marcos de error "aceptables" considerar que puede haber ciertas diferencias entre ambas. Pueden manifestarse por diferentes motivos:

- Puede ocurrir, por ejemplo, que ciertas escuelas seleccionadas no participen (puede ser por problemas de logística, decisiones de la institución, u otras razones). PISA, en estos casos, prevé estrategias de reemplazo, para evitar que estas ausencias impacten en la cobertura.
- Otro ejemplo puede ser que haya estudiantes que se ausentan al momento de la evaluación, o que ya no asisten a la escuela. En el operativo PISA, el aplicador asiste a la escuela con un listado de

estudiantes seleccionados para evaluar, que es confeccionado unos meses antes. La diferencia temporal entre la confección del listado y la selección final de estudiantes puede ocasionar que haya estudiantes que por distintas razones no asistan a la escuela. En el cálculo de los factores de ponderación de la muestra, PISA considera estas situaciones para ajustar los pesos asignados a los estudiantes evaluados (expansión del peso base), de modo que la estimación global compense estas ausencias.

- También puede generar diferencia la imprecisión del dato de estudiantes que se considera en el marco muestral. Este es un motivo bastante común, debido a que los marcos muestrales se elaboran generalmente en el período escolar anterior a la implementación de PISA. Por lo tanto, el total de estudiantes por escuela no refiere al año de aplicación, sino que puede tener uno o dos años de antigüedad. Las escuelas seleccionadas para evaluar pueden tener menos cantidad de estudiantes que los que se esperaba encontrar según estos datos desactualizados. Los países con sistemas de información más débiles suelen tener mayores dificultades para elaborar estas estimaciones.

Esos son sólo algunos ejemplos puntuales de razones por las cuales la muestra ponderada puede ser menor al marco muestral. Sin embargo, la representatividad de la muestra comienza a ponerse en juego cuando estos problemas se intensifican. Cuanto mayor es la brecha entre la muestra ponderada y la población objetivo deseada, surgen interrogantes clave: ¿en qué medida se ve afectada la representatividad de los resultados? ¿Cuáles son los límites para considerar a la muestra como válida y comparable? ¿Las comparaciones a lo largo del tiempo se ven afectadas por estos bajos niveles de cobertura?

Por otra parte, todos estos problemas de representatividad no sólo ponen en jaque la validez de las estimaciones nacionales, también tienen impacto en el análisis de las brechas de desigualdad y del abordaje en torno a los factores asociados.

En las publicaciones de PISA no existen criterios claros para establecer en qué situaciones, o bajo qué condiciones, estas diferencias entre la muestra ponderada y la población objetivo definida permiten realizar estimaciones representativas y comparables. Este índice se utiliza con fines de diagnósticos y no está directamente relacionados con los estándares técnicos. Cuando asume valores bajos, se suele solicitar a los países justificaciones, ya que podrían vincularse al incumplimiento de algún estándar. Sin embargo, en los casos presentados no existe evidencia en las publicaciones que dé cuenta de las razones por las cuales las muestras difieren tanto de la población objetivo.

La identificación de países que, como el caso de Argentina en PISA 2015, muestran valores bajos en el índice de cobertura 4, alerta sobre la necesidad de precisar los estándares mínimos de diferencia entre muestra ponderada y población objetivo definida. El interrogante permanece sin resolver: ¿en qué medida estos bajos niveles de cobertura ponen en juego la representatividad nacional de los resultados, y su comparabilidad en el tiempo?