**La segunda dimensión de la calidad educativa:**

**Un estudio comparado en siete países de América Latina**

**The Second Dimension of Educational Quality:**

**A Comparative Study in Seven Countries in Latin America**

**Axel Rivas (\*) y Martín Scasso (\*\*)**

**(\*) Universidad de San Andrés, Argentina**

**(\*\*) Fundación Quantitas**

**Resumen**

INTRODUCCIÓN. Las evaluaciones de la calidad educativa permitieron indagar en los últimos 20 años la comparación de los sistemas educativos. Los resultados de las pruebas PISA son los más utilizados como referencia. En América Latina siete países participan de manera recurrente desde el año 2000: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay.

MÉTODO. Se analiza la evolución de estos países en la prueba PISA y en otras evaluaciones nacionales e internacionales. Se analizan los objetivos, sesgos metodológicos y resultados comparados de estas evaluaciones.

RESULTADOS. Las tendencias que muestran los resultados de las distintas evaluaciones de la calidad tienen contradicciones entre sí, incluso con variaciones basadas en problemas metodológicos poco explorados hasta ahora. Las pruebas PISA introdujeron cambios en la edición 2015 que modifica los resultados de las ediciones anteriores y hace incomparables los resultados de las primeras ediciones (2000 y 2003). Se abren interrogantes a partir de la constatación de que los sistemas educativos que parecían haber mejorado en realidad no lo han hecho o lo han logrado de manera más reducida. Se presenta una mirada comparada de los resultados de las evaluaciones de los siete países analizados, constatando que hubo casos de mejora como el de Perú desde un punto de partida más bajo y el de Chile en la década del 2000, alcanzando el punto más alto entre los países analizados.

DISCUSION. El estudio presenta la necesidad de discutir una “segunda dimensión de la calidad educativa”, que requiere ir más allá de las lecturas simplificadas de los resultados de las evaluaciones, tomando diversas fuentes e indicadores educativos de manera complementaria. Esa segunda dimensión funciona como un filtro epistemológico para comprender la evolución de los sistemas educativos en América Latina y evitar lecturas equivocadas o sesgadas de la calidad educativa.

**Palabras claves:** Evaluación educativa, Política educativa, Calidad educativa, Análisis comparado, Re-escalamiento.

**Introducción**

Analizar los aprendizajes de los alumnos en los sistemas educativos de América Latina es una tarea que cobró un significado cuantitativo en los últimos 30 años. En los años noventa comenzaron a surgir en casi todos los países los sistemas de evaluación de la calidad educativa, que tenían en su primera etapa la prioridad de diagnosticar el estado de los aprendizajes en áreas como Lengua y Matemática (Wolf, 2004; Ferrer, 2006; Swaffield y Thomas, 2016).

En 1997 se lanzó la primera prueba comparada del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la UNESCO para el nivel primario. En 2000 comenzó el Programa Internacional para la Evaluación de Estudiantes (PISA) de la Organización para la Cooperación y el Desarrollo Económico (OCDE), que atrajo la participación regular de siete países de la región en sus sucesivas ediciones: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Este artículo parte de la prueba PISA y consolida una mirada integral sobre la calidad educativa en estos siete países, a partir de las posibilidades de comparar sus resultados en el tiempo.

La aparición de las evaluaciones de la calidad nos permite entender de manera sistemática qué aprenden los alumnos (Morris, 2011). Dan luz sobre la caja negra de las aulas y ayudan a convertir a la educación en un derecho social que puede ser monitoreado. Las pruebas permiten diagnosticar problemas no conocidos de los sistemas educativos con precisión y consistencia (Lüdtke, Robitzsch, Trautwein, Kreuter y Ihme Marten, 2007). Sus cuestionarios complementarios sirven para medir las desigualdades en los resultados y los factores asociados a la calidad educativa. Esto abre numerosas avenidas de indagación científica para comprender qué puede hacerse para mejorar la educación a escala sistémica. En los años 2000 las evaluaciones de muchos países de la región pasaron a tener un rol de alto impacto en las escuelas. Ya no sólo se trataba de medir al sistema sino también comprender qué pasaba en cada escuela. Esto permitió iluminar la situación de los aprendizajes construyendo mecanismos de apoyo a las escuelas con más dificultades. También en algunos países se combinaron medidas de presión sobre las escuelas, con publicación de resultados o incluso beneficios económicos condicionados a los rendimientos (Ferrer y Fiszbein, 2015).

Las evaluaciones trajeron consigo distintos tipos de críticas. Algunos concentraron su mirada en el reduccionismo de las pruebas, que cuanto más alto impacto tienen más reducen el currículo a aquello que es evaluado, restando importancia al resto de la enseñanza (Berliner, 2011). Otros autores indicaron que las pruebas producen una presión excesiva sobre las escuelas (Fuller, Henne y Hannum, 2008) y una deslegitimación constante de los docentes, que en ciertos casos es utilizada para agitar una emergencia educativa y aplicar reformas neoliberales de privatización del sistema (Ball, 2003; Taubman, 2009). Las evaluaciones internacionales desataron también una ola de críticas basadas en el efecto colonizador de PISA, señalando su capacidad de incidir en el currículo nacional de los países a partir de criterios económicos (Grek, Lawn y Ozga, 2009). Varios autores analizaron cómo las pruebas PISA habían cobrado un carácter cada vez más decisivo en la agenda de política educativa de los países participantes (Pons, 2017; Lindblad, Pettersson y Popkewitz, 2015; Fischman, Marcetti Topper, Silova, Goebel y Holloway, 2018; Rautalin, Alasuutari y Vento, 2018; Lewis y Lingard, 2015; Niemann, Martens y Teltemann, 2017; Meyer y Benavot, 2013). Como resaltaron algunos de estos trabajos, las pruebas PISA abrieron una grieta entre un uso acrítico de las evaluaciones, que toma sus resultados de manera literal, y una crítica extrema que invita a dejar de medir los sistemas de manera estandarizada. Este trabajo se sitúa en una perspectiva intermedia, buscando un uso con criterio de las evaluaciones disponibles para medir la calidad educativa. Otros estudios ya advirtieron sobre los problemas de comparabilidad entre países de pruebas como PISA (Eivers, 2010; Kreiner y Christensen, 2014; Cettin, 2010; Chen y Jiao, 2014) y las irregularidades que presentan distintos problemas metodológicos (Hopfenbeck et al., 2017; Lindblad, Pettersson y Popkewitz, 2015; Fernández-Cano, 2016). Algunos trabajos previos detectaron fallas en la comparabilidad inter-temporal de las pruebas PISA (Carnoy y Rothstein, 2013; Freitas, Catela Nunes, Balcão Reis, Seabra y Ferro, 2016; Cosgrove, 2015; Eivers, 2010). En la región, el caso de Brasil ya fue revisado con una crítica metodológica a la comparabilidad de los resultados (Klein, 2011; Carnoy, Khavenson, Fonseca, Costa y Marotta, 2015). Estos estudios son antecedentes directos del presente artículo y nos muestran la necesidad de analizar rigurosamente las evaluaciones de la calidad para entender su validez, tanto en términos de lo que miden como de las posibles extrapolaciones políticas.

El objetivo de este artículo es construir una doble mirada indisociable entre la primera dimensión de la calidad educativa, definida por los resultados de las evaluaciones estandarizadas, y una segunda dimensión que toma distancia crítica revisando su metodología y contrastando distintas fuentes. La hipótesis central es que hoy no se pueden leer los resultados de las evaluaciones sin un cuidado metodológico básico sobre su construcción metodológica.

**Método: una comparación sistemática de evaluaciones de la calidad educativa**

Para cumplir este objetivo se recurrió a dos estrategias metodológicas. La primera se centró en el análisis de los resultados arrojados por las sucesivas ediciones de la prueba PISA. A partir de la edición 2015 se constató un cambio metodológico que tuvo importantes consecuencias sobre la comparabilidad inter-temporal de los países de América Latina, escasamente caracterizado en el informe de la Organización para la Cooperación y el Desarrollo Económicos​ (OECD, 2016a). Las hipótesis de investigación que se formularon para explicar por qué este cambio afectaba sólo a la región latinoamericana permitieron enfocar la mirada en el modo en que los estudiantes responden u omiten ítems en las pruebas PISA. Fue necesario analizar las respuestas a nivel de ítem, por alumno, en todas las pruebas de todos los países participantes de los ciclos 2000 – 2015.

La segunda estrategia metodológica consistió en ampliar la mirada y sistematizar los resultados de otras evaluaciones internacionales, especialmente de las pruebas SERCE y TERCE de la UNESCO y de las evaluaciones nacionales de los siete países estudiados: ONE/Aprender en Argentina, SAEB en Brasil, SIMCE en Chile, SABER en Colombia, ENLACE en México y ECE en Perú1. Estas evaluaciones tienen rasgos distintos que no permiten su comparación directa sin comprender sus objetivos y estrategias metodológicas.

Nuestro trabajo consistió en sistematizar las evaluaciones de los países, considerando el modo en que estandarizan la dispersión de cada escala de puntaje, tomando en cuenta las tendencias centrales y las desviaciones estándar definidas. Las evaluaciones de cada país tienen objetivos distintos a las internacionales y la comparación no indica que deberían obtener resultados idénticos. Incluso, en algunos casos no es clara la validez de la comparación en el tiempo de la misma evaluación nacional en todas sus aplicaciones. Lo que busca este trabajo es dar una mirada más integral al análisis de la calidad educativa, tomando en cuenta todas las fuentes disponibles sobre los resultados de siete países en distintas evaluaciones. Así se busca abordar una segunda capa de la calidad educativa, que controla por vía de la multiplicación de las fuentes lo que sabemos sobre los resultados de aprendizaje en los siete países analizados.

No existen antecedentes de trabajos similares en América Latina. Otros estudios analizaron las pruebas PISA con perspectiva iberoamericana (Grupo Iberoamericano de PISA, 2010; OECD, 2018) o latinoamericana (Bos, Elías, Vegas y Zoido, 2016), existen análisis profundos sobre las pruebas de la UNESCO (LLECE, 2014), miradas integradas de ambas evaluaciones (Hanushek y Woessmann, 2012) y trabajos que en cada país han consolidado visiones sobre los resultados de las evaluaciones (Instituto Colombiano para la Evaluación de la Educación, 2017; Ministério da Educação do Brasil, 2016). El presente artículo es original en la síntesis que realiza de estas distintas fuentes para obtener una mirada panorámica de la calidad educativa en América Latina.

**Resultados**

**América Latina en las pruebas PISA: el problema de la comparabilidad**

La prueba PISA de la OCDE se aplicó por primera vez en el año 2000 y evalúa a alumnos de 15 años en Lengua, Matemática y Ciencias. En 2015 participaron 72 países, 35 de la OCDE y nueve de América Latina, de los cuales siete habían venido participando en varias ediciones desde el año 2000: Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay. Sobre estos siete países se concentrará el análisis de los resultados, dado que es posible compararlos en el tiempo.

En la última edición de la prueba PISA se incorporaron cambios metodológicos que buscaron mejorar la precisión de las mediciones. Un cambio en particular tuvo un impacto destacado en varios países de América Latina, pero no en otros países participantes del resto del mundo. Las ediciones de PISA de 2000 a 2012 contabilizaban a las pruebas entregadas de manera incompleta por los alumnos como válidas y les asignaban el puntaje más bajo a las preguntas no respondidas. En 2015 se decidió cambiar el criterio y considerar no válida esta falta de respuesta (OECD, 2016a).

El cambio en la valoración de las pruebas con ítems no completados (*non-reached ítems*) del año 2015 puso en evidencia que este problema estaba únicamente presente en países de América Latina. En la región el porcentaje de ítems no completados era del 11,9% en 2000 y fue disminuyendo hasta llegar al 1,9% en 2015, mientras en el resto de los países fue mucho más bajo desde un comienzo.

El cambio metodológico presenta tres problemas para la comparabilidad inter-temporal de los resultados en América Latina. El primero es que este cambio afectó los resultados de algunos países. La tabla 1 presenta los resultados originales y re-escalados para los países participantes de América Latina2. Cabe resaltar que los resultados re-escalados son los correctos para toda comparación inter-temporal, dado que se basan en la equiparación metodológica de los ítems no completados.

En la tabla 1 se observa que las mejoras notables de Colombia entre 2012 y 2015 fueron sólo el efecto del cambio de metodología. En Uruguay los resultados originales mostraban cierta estabilidad, con una sustantiva mejora en la última aplicación, pero con los resultados calculados con la metodología de PISA 2015, el escenario que se perfila es inverso. En el caso de Argentina, que se presenta sólo hasta 20123, el escenario de mejora de los resultados con los datos originales tiende a reducirse. En Perú el cambio de metodología también incide, pero la tendencia de mejora se mantiene en ambos escenarios, con menor intensidad en los resultados re-escalados.

Chile es un caso de incidencia distinta en el cambio de la metodología: con los resultados originales, Chile mostraba una tendencia de mejora en los primeros años (2000 a 2006), que ahora ya no puede ser comparado con las pruebas de los años siguientes. Brasil y México son los países menos afectados por estos cambios: en promedio, los resultados de las tres áreas muestran valores similares en el periodo 2006 – 2015.

Tabla 1. Resultados PISA en puntajes originales (2000-2015) y re-escalados (2006-2015), por área.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Lectura** | **Matemática** | **Ciencias** |
| **Argentina (\*)** |  |  |  |
| **Brasil** |  |  |  |
| **Chile** |  |  |  |
| **Colombia** |  |  |  |
| **México** |  |  |  |
| **Perú (i)** |  |  |  |
| **Uruguay** |  |  |  |



(\*) Los datos de Argentina fueron incluidos hasta 2012, debido que los resultados 2015 no se consideran comparables a causa de la baja cantidad de estudiantes representados en la muestra.

Fuente: procesamientos propios en base a PISA 2015.

El segundo problema es comunicacional: la OCDE hizo el ejercicio de re-escalar los resultados originales con el nuevo criterio de corrección de 2015 pero no los presentó en sus informes centrales, sino apenas en los anexos (OECD, 2016a) y en estudios más recientes sobre los resultados en Iberoamérica (OECD, 2018). Incluso, en los propios informes centrales se toman como casos de éxito a países de América Latina con los resultados originales que luego son corregidos en los anexos. Esto ocurre con Colombia en el informe central de PISA, mencionado como caso de éxito (OECD, 2016a, p. 3), y en el informe del país, donde no se hace ninguna mención al ejercicio que corrige los resultados y se resalta la mejora en los resultados (OECD, 2016b, p. 1).

Tal es así que el propio ex presidente de Colombia festejó los avances del país en PISA cuando se publicaron los resultados en diciembre de 20164 y los informes nacionales también tomaron los resultados originales sin hacer referencia a las correcciones (Instituto Colombiano para la Evaluación de la Educación, 2017). Es muy probable que en varios países analizados aquí se estén asumiendo resultados de mejora que no fueron tales por esta confusión comunicacional.

El tercer problema es la ruptura de la comparabilidad de la prueba PISA para toda la serie histórica. El ejercicio de re-escalamiento de los resultados con la nueva metodología de 2015 fue realizado por PISA para las ediciones 2012, 2009 y 2006. Los resultados para las ediciones 2000 y 2003 ya no son comparables con esta nueva metodología. Justamente el año 2000, que sitúa el punto de partida de estas mejoras, tuvo tasas de no respuesta muy altas en todos los países de América Latina. ¿Fueron las mejoras posteriores el efecto de mejoras en los aprendizajes o sólo en la tasa de respuesta?

No podemos contestar esta pregunta empíricamente, pero sí es posible formular una hipótesis: a medida que las pruebas PISA tuvieron más alto impacto político, creció la presión para que los alumnos contesten con mayor esfuerzo. Este fenómeno se vincularía con la generalización de una cultura de la evaluación (Jodouin, 2014), con la expansión de las evaluaciones nacionales (Wolff, 2004) y de políticas que asignaron mayor impacto a los resultados de las pruebas estandarizadas, y con un mayor impacto mediático y político de la publicación de los resultados de cada edición de PISA. Estas hipótesis provienen de un estudio paralelo donde se consultó a más de 130 actores políticos de la educación en los siete países estudiados (Rivas, 2015).

Esto generó una paradoja: el aumento de la tasa de respuesta de los alumnos hizo la prueba más válida (disminuyendo las tasas de no respuesta), pero perdió comparabilidad en el tiempo, dado que habían operado criterios distintos de esfuerzo para la prueba en las distintas ediciones.

La literatura sobre el impacto de la motivación de los alumnos en las pruebas (*Test-Taking Motivation*) señala la necesidad de diferenciar la motivación (*will*) de las habilidades (*skill*) (Eklof, 2010). Diversos estudios muestran que en las pruebas de bajo impacto (*low-stakes*) los alumnos realizan un menor esfuerzo (Wise y DeMars, 2005; Pintrich y DeGroot, 1990; Mislevy, 1995; Wolf y Smith, 1995; Finn, 2015). Esto varía mucho según contextos culturales (Wuttke, 2007; Gneezy et al., 2017). Así se comprueba en nuestro estudio: en América Latina la falta de motivación para realizar las pruebas PISA con el mayor esfuerzo parece haber sido más alta en las primeras ediciones que en países de otras regiones.

La disminución de la tasa de no respuesta puede haber sido el efecto de la creciente presión política que tuvo PISA, con la amplia difusión mediática de sus resultados (Ravela, 2003; Pons, 2017). Esta presión pudo haberse trasladado a una mayor preparación para la prueba en la muestra de escuelas que recibe cada país seis meses antes de la aplicación. A su vez, la mejora en las tasas de respuesta puede deberse a un avance en la cultura de la evaluación en la región, con la implementación ya acostumbrada en muchos países de pruebas nacionales (Solano-Flores y Milbourn, 2016).

**Ensanchar la mirada: comparación de resultados en otras evaluaciones**

Las pruebas PISA dejan más preguntas que respuestas sobre la comparabilidad inter-temporal de los resultados en cada país. Para los fines de este artículo, la pregunta acerca de la evolución de la calidad educativa en América Latina requiere ampliar la mirada y comparar los resultados de otras evaluaciones nacionales e internacionales.

Las tablas 2 y 3 comparan todas las pruebas analizadas de los países, que se suman a la tabla 1 con los resultados de PISA. Para simplificar al máximo el análisis se decidió tomar como parámetro de referencia el periodo 2006-2015 con los resultados re-escalados de PISA en el nivel secundario y el período 2007-2013 de la prueba SERCE-TERCE de la UNESCO para el nivel primario.

Tabla 2: Resultados de evaluaciones internacionales (PISA – resultados reescalados –, SERCE/TERCE), por área. Países seleccionados de América Latina. 2006 – 2015

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PAÍSES | SERCE/TERCE  (6to grado)\*\*\* | | | | | | PISA – resultados reescalados  (estudiantes 15 años) | | | | | |
| Lectura | | | Matemática | | | Lectura | | | Matemática | | |
| 2006 | 2013 | *Dif.* | *2006* | 2013 | *Dif.* | 2006 | 2015 | *Dif.* | 2006 | 2015 | *Dif.* |
| Argentina\* | 506 | 509 | *2* | *513* | 530 | *17* | 392 | 400 | *8* | 397 | 389 | *-8* |
| Brasil | 520 | 524 | *4* | *499* | 520 | *20* | 399 | 407 | *8* | 371 | 377 | *6* |
| Chile | 546 | 557 | *11* | *517* | 581 | *63* | 450 | 459 | *9* | 417 | 423 | *6* |
| Colombia | 515 | 526 | *11* | *493* | 515 | *22* | 421 | 425 | *4* | 385 | 390 | *5* |
| México | 530 | 529 | *-1* | *542* | 566 | *24* | 417 | 423 | *6* | 407 | 408 | *1* |
| Perú\*\* | 476 | 505 | *29* | *490* | 527 | *37* | 383 | 398 | *15* | 371 | 390 | *19* |
| Uruguay | 542 | 532 | *-10* | *578* | 567 | *-12* | 427 | 437 | *10* | 437 | 418 | *-19* |

Las diferencias que son estadísticamente significativas, considerando un nivel de significación de 0,05, aparecen en negrita.

\* Los resultados de Argentina de PISA corresponden al período 2006 – 2012, debido a la baja cantidad de estudiantes representados en la muestra de 2015

\*\* Los resultados de Perú de PISA corresponden al período 2009 – 2015, debido a que el país no participó en la edición 2006.

\*\*\* Grado que corresponde teóricamente a la población de 11 años de edad. Debido a los altos niveles de sobreedad existentes en estos países, también asisten a estos grados estudiantes de edad superior

Fuente: Elaboración propia en base a datos de OREALC/UNESCO SERCE Y TERCE, y OECD-PISA.

Tabla 3: Resultados de las evaluaciones nacionales, por área. Años de estudio seleccionados. Países de América Latina. 2006 – 2015

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Argentina – ONE** | | *% de estudiantes con logro medios y altos* | | |
| **2007** | **2010** | **2013** |
| 6º primariaIV | Lengua | 66,3% | 72,3% | 72,0% |
| Matemática | 63,8% | 64,3% | 64,4% |
| 5º secundaria VII | Lengua | 70,8% | 75,6% | 75,6% |
| Matemática | 35,0% | 44,1% | 45,8% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Brasil – SAEB** | | *Puntaje promedio estandarizado. Media = 250 y desv. est. = 50* | | | | |
| **2007** | **2009** | **2011** | **2013** | **2015** |
| 5º Enseñanza fundamentalIII | Lengua | 176 | 184 | 191 | 196 | 196 |
| Matemática | 193 | 204 | 210 | 211 | 219 |
| 9º Enseñanza FundamentalV | Lengua | 235 | 244 | 245 | 245 | 252 |
| Matemática | 247 | 249 | 253 | 252 | 256 |
| 3ª serie Enseñanza mediaVII | Lengua | 261 | 265 | 269 | 269 | 269 |
| Matemática | 402 | 406 | 409 | 409 | 409 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Chile – SIMCE** | | *Puntaje promedio estandarizado. Media = 250 y desv. est. = 50* | | | | | | | | | |
| **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** |
| 4º educación básicaII | Lengua | 253 | 254 | 260 | 262 | 271 | 267 | 267 | 264 | 264 | 265 |
| Matemática | 248 | 246 | 247 | 253 | 253 | 259 | 261 | 256 | 256 | 260 |
| 2º educación mediaVI | Lengua | 254 |  | 255 |  | 259 |  | 259 | 254 | 252 | 247 |
| Matemática | 252 |  | 250 |  | 256 |  | 265 | 267 | 265 | 262 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Colombia – SABER** | | *Puntaje promedio estandarizado. Media = 300 y desv. est. = 80* | | | | | | | | | |
| **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** |
| 5º básicoIII | Lengua |  |  |  | 301 |  |  | 302 | 303 | 297 | 297 |
| Matemática |  |  |  | 301 |  |  | 294 | 299 | 292 | 301 |
| 9º básicoV | Lengua |  |  |  | 301 |  |  | 307 | 300 | 297 | 295 |
| Matemática |  |  |  | 302 |  |  | 302 | 300 | 296 | 296 |
|  |  | *Puntaje promedio estandarizado. Media = 50 y desv. est. = 10* | | | | | | | | | |
| 2do nivel medioVI | Lengua | 48 | 47 | 45 | 49 | 50 | 46 | 47 | 47 | 50 | 50 |
| Matemática | 46 | 45 | 44 | 45 | 50 | 46 | 50 | 45 | 50 | 50 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **México – ENLACE** | | *Puntaje promedio estandarizado. Media = 500 y desv. est. = 100* | | | | | | | | |
| **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** |
| 5º básicoIII | Lengua | 500 | 509 | 512 | 514 | 545 | 549 | 548 | 550 |  |
| Matemática | 500 | 510 | 517 | 527 | 547 | 551 | 571 | 588 |  |
|  | | *% agrupado de estudiantes con logro elemental, bueno y excelente* | | | | | | | | |
| 3º secundariaV | Lengua |  |  | 88% | 83% | 88% | 85% | 86% | 84% | 81% |
| Matemática |  |  | 53% | 54% | 59% | 65% | 70% | 72% | 73% |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Perú –ECE** | | *Puntaje promedio estandarizado. Media = 500 y desv. est. = 100* | | | | | | | | |
| **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** |
| 2º grado primariaI | Lengua | 500 | 503 | 524 | 536 | 535 | 541 | 545 | 568 | 592 |
| Matemática | 500 | 512 | 520 | 519 | 519 | 523 | 526 | 560 | 575 |

Fuente: elaboración propia en base a resultados de ONE 2007- 2013, SAEB 1999-2015, SIMCE 2001-2016, SABER 2005-2016, ENLACE 2008-2014 y ECE 2007-2015.

I, II, III, IV, V, VI, VII A continuación se detallan las edades teóricas que corresponden a cada año de estudio. Es relevante considerar que, debido a los altos niveles de sobreedad existentes en estos países, también asisten a estos grados estudiantes de edades superiores. I: 7 años, II: 9 años, III: 10 años, IV: 11 años, V: 14 años, VI: 16 años, VII: 17 años. En anexo se presenta una tabla comparativa de estructuras de niveles. Al analizar las tablas se pueden encontrar trayectorias muy variadas de los países. **Perú** muestra mejoras en todas las pruebas analizadas desde un punto de partida muy bajo hasta llegar más cerca del promedio regional, tanto en primaria como en secundaria. La prueba nacional de 2do grado (Tabla 3) perfila una mejora muy marcada. Entre 2007 y 2015 en lectura el puntaje se incrementa 92 puntos, y en Matemática 75 (escala con desviación estándar de 100 puntos). Estos resultados también se reflejan en las evaluaciones SERCE/TERCE (Tabla 2).

También se destaca la mejora en las evaluaciones de PISA (Tabla 2). Es el único país que muestra mejoras estadísticamente significativas con los resultados re-escalados de 2015. Esta mejora se logró en un contexto de grandes avances en la inclusión y la trayectoria de los alumnos (Rivas, 2015). Con 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 el nivel primario, manteniéndose por encima de los demás países analizados. Las mejoras en la prueba de la UNESCO se destacan en Matemática, mientras en SIMCE fueron mucho más leves.

En el nivel secundario, los resultados de PISA (Tabla 2) muestran cierto estancamiento en ambas áreas, con una mejora no significativa. Cabe señalar que las mejoras en PISA se dieron sobre todo en el ciclo 2000 a 2006, que ya no puede ser comparado con las pruebas re-escaladas de 2015, dejando abierto el interrogante si se trató de una mejora en los aprendizajes o en la tasa de respuesta de los alumnos. Los resultados en las pruebas nacionales en el nivel medio (Tabla 3) son en cierto sentido contradictorios, con mejoras en Matemática y cierto retroceso en lectura en los años recientes. De todas formas, Chile se mantiene por encima del resto de los países en los resultados de PISA.

**Brasil** muestra una marcada diferencia en sus avances en la calidad educativa en el nivel primario y medio. En el nivel primario el país mejoró en las pruebas internacionales de la UNESCO en Matemática (Tabla 2), con un incremento de 20 puntos, pero no así en lectura. Las pruebas nacionales (Tabla 3) también indican un ciclo de mejora desde 2007 hasta 2015, con incrementos importantes en ambas áreas (20 en lectura y 26 en Matemática, en una escala con una desviación estándar de 50 puntos).

En el nivel medio, la prueba SAEB (Tabla 3) muestra mejoras leves en el noveno año, no así en el último año de secundaria. Mientras tanto, la prueba PISA (Tabla 2) no muestra avances significativos: la mejora de Brasil se identifica más en la primera etapa, entre 2000 y 2006 (período que ya no se puede comparar con los resultados re-escalados). 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 (Rivas, 2015).

**México** también tiende a un comportamiento divergente en el nivel secundario y convergente en primaria. Lo más notable es la mejora en Matemática en el nivel primario, tanto en las pruebas internacionales (incremento de 24 puntos) (Tabla 2) como, especialmente, en las nacionales (Tabla 3), donde crece 88 puntos en una escala cuya desviación estándar es de 100. Lengua muestra, para la prueba nacional, una mejora significativa pero más leve, mientras que en las pruebas UNESCO se mantiene estable.

En PISA el país oscila con vaivenes que tienden a un escenario de estabilidad (Tabla 2). En las pruebas nacionales (Tabla 3) se observan comportamientos opuestos: mejoras muy marcadas en Matemática, mientras que en lectura hay una leve tendencia a la caída. También México logró un incremento de las tasas de acceso al nivel medio, desde un punto de partida muy bajo.

En **Colombia**, los resultados muestran estabilidad en ambos niveles educativos. Las pruebas de UNESCO de primaria (Tabla 2) presentan una leve tendencia a la mejora, que es estadísticamente significativa sólo en Matemática. Las pruebas nacionales (Tabla 3), por el contrario, se muestran invariantes en la comparación. En las pruebas internacionales (Tabla 2) de PISA (luego de ser re-escaladas para ser comparables en el tiempo) y en las pruebas nacionales SABER (Tabla 2), prácticamente no hay variaciones. La serie de tiempo completa muestra vaivenes difíciles de explicar. Lo que sí se destaca es un gran avance del país en el aumento de las tasas de escolaridad y reducción de la sobreedad en el nivel secundario (Rivas, 2015).

En **Uruguay** se observa un ciclo de estabilidad en los resultados en el nivel primario en las pruebas internacionales y una tendencia divergente en secundaria. No existen pruebas nacionales para contrastar estos avances. Los resultados de las pruebas UNESCO (Tabla 2) posicionan al país como el único que no ha alcanzado mejoras en ninguna de las dos áreas analizadas. En PISA, Uruguay muestra dos tendencias opuestas: mejora leve en lectura, de 10 puntos de diferencia, que no alcanza significancia estadística, y caída en Matemática de cerca de 20 puntos.

En **Argentina** hay incrementos en el nivel primario tanto en las pruebas nacionales como internacionales, aunque en el primer caso la mejora se observa más intensamente en lectura, mientras que en el segundo sólo en Matemática. En las pruebas nacionales (Tabla 3) no hay posibilidad de establecer la significancia estadística de estas diferencias, y en las pruebas SERCE-TERCE (Tabla 2) la mejora es significativa sólo en Matemática (17 puntos). En secundaria los escenarios también son divergentes: en las pruebas nacionales (Tabla 3) se constatan mejoras, más intensas en Matemática que en Lengua. En cambio, en las pruebas PISA (Tabla 2) hay estabilidad hasta el año 2012, mientras en la prueba 2015 fue anulada su participación por un problema en el marco muestral.

Esta breve síntesis muestra algunas de las contradicciones entre distintos tipos de pruebas nacionales e internacionales. La mirada más específica a cada variación de un año a otro de cada prueba abre más contradicciones, que no han sido planteadas aquí por una cuestión de espacio. Las pruebas analizadas tienen criterios distintos, pero resulta llamativo que tengan en tantos casos resultados divergentes.

**Discusión**

¿Qué países mejoraron la calidad de los aprendizajes en América Latina? ¿Cómo sabemos si mejoraron? Las respuestas a la primera pregunta parecen simples. La aparición de las pruebas estandarizadas como PISA favorece una perspectiva comparada de los sistemas educativos entre países y en su evolución temporal. Bastaría leer las evaluaciones internacionales para comparar los resultados. Algunos analistas y tomadores de decisiones recurren a este novedoso recurso para abreviar al máximo la compleja definición de “calidad educativa”. Esto constituye la primera dimensión de la calidad educativa.

Este artículo muestra que las respuestas a la segunda pregunta no son tan sencillas. Comprender qué miden las evaluaciones, qué tan fiables son y cuáles son sus riesgos metodológicos nos lleva a la segunda dimensión de la calidad educativa. En este artículo mostramos que no es posible entender qué países mejoraron en América Latina sin saber realmente qué miden las evaluaciones.

El artículo está basado en dos tipos de evidencias empíricas. La primera evidencia empírica está centrada en la prueba PISA de la OECD. En base a la reconstrucción de las tasas de respuesta de los alumnos de los países de América Latina que participan de PISA, hemos hallado que existen serios problemas para garantizar la validez metodológica de las comparaciones inter-temporales de los resultados.

Un segundo tipo de evidencia empírica permite ensanchar la mirada a partir de la comparación sistemática de distintos tipos de evaluaciones de la calidad educativa aplicadas en América Latina. El estudio comparado de pruebas nacionales e internacionales nos permite mostrar múltiples contradicciones entre los resultados.

En este trabajo se presentan evidencias concretas que contestan la pregunta acerca de qué países mejoraron la calidad de los aprendizajes en América Latina con una respuesta frustrante pero realista: no lo sabemos de manera concluyente. Necesitamos mirar distintas fuentes de evaluación de los aprendizajes, desarrollar estudios longitudinales de valor agregado, y de este modo ampliar las posibilidades de abordaje para aproximarnos a respuestas siempre parciales pero más robustas.

Esto es importante porque las lecturas rápidas, sesgadas o no, de los medios de comunicación y de muchos analistas, pueden llegar a ser equivocadas y elaborar diagnósticos mal enfocados en base a información errónea.

Para entender qué países mejoraron la calidad educativa debemos conocer los problemas metodológicos de las pruebas y ensanchar la mirada aprovechando la multiplicación de evaluaciones de estos últimos años. Esto es apenas la punta del iceberg. Lo que miden las pruebas estandarizadas es una foto parcial de lo que ocurre en la enorme complejidad de un sistema educativo.

Como indican las miradas más profundas de la educación comparada:

“PISA es una brillante forma de gran-ciencia-social para mapear resultados, pero de ninguna manera antropológica, histórica o cultural es trabajo ´comparado´. Sin embargo, es una contribución a la educación comparada porque establece nuevos laberintos antropológicos, históricos y culturales acerca de qué significan esos resultados” (Cowen, 2011, p. 261).

PISA y otras evaluaciones son fuentes necesarias, pero no determinantes. Para leer sus resultados hay que tomar precauciones metodológicas, comparar diversas fuentes y seguir las series de resultados durante varias ediciones en el tiempo, sabiendo que en todas las pruebas hay cambios metodológicos o de contextos de aplicación. La estabilización de los resultados en varias pruebas nos permite establecer mayores niveles de fiabilidad y confianza en la validez interpretativa de las pruebas. Así podemos ver que casos como Perú han logrado mejoras notables en diversos registros, mientras Chile parece haber logrado mejoras en la primera década del siglo XXI y ubicarse por encima del resto de los países participantes de las pruebas internacionales en América Latina.

Para seguir comprendiendo mejor estos procesos es necesario apelar a estudios sobre la cultura que rodea a los sistemas educativos (Alexander, 2001; Carnoy, 2007) y hacia evaluaciones más integrales. Es importante analizar las características del currículo de cada país y sus tradiciones culturales y pedagógicas. Esta es la tercera dimensión de la calidad educativa, que incorpora diversas perspectivas metodológicas y realiza análisis históricos, culturales, sociológicos y pedagógicos de los sistemas educativos, buscando comprender en qué contextos aprenden los alumnos.

**Anexo – estructura de niveles y años de estudio**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Año de estudio | Edad teórica | Argentina | Brasil | Chile | Colombia | México | Perú | Uruguay |
| 1º | 6 años | Educación Primaria | Enseñanza Fundamental | Educación básica | Educación Primaria | Educación básica | Educación básica | Educación Primaria |
| 2º | 7 años |
| 3º | 8 años |
| 4º | 9 años |
| 5º | 10 años |
| 6º | 11 años | Educación Secundaria |
| 7º | 12 años | Educación Secundaria Básica | Educación Secundaria | Educación Secundaria | Educación Secundaria Ciclo Básico |
| 8º | 13 años |
| 9º | 14 años | Educación Media |
| 10º | 15 años | Educación Secundaria Orientada | Enseñanza Media | Educación Media | Educación Media Superior | Educación Secundaria Bachillerato |
| 11º | 16 años |
| 12º | 17 años |  |  |

Fuente: Elaboración propia

**Notas**

1 Sólo Uruguay no tuvo evaluaciones nacionales durante este período.

2 La OECD incluye en el informe de PISA 2015 (OECD, 2016a) los resultados de las pruebas 2006 a 2012 calculados en base a la metodología utilizada en 2015 (referidos en este documento como “resultados re-escalados”. Utiliza esta información para realizar estimaciones del impacto del cambio de metodología. Sin embargo, estas estimaciones son globales y no se detiene a analizar el caso de América Latina.

3 Los resultados de Argentina de PISA corresponden al período 2006 – 2012, debido a que la baja cantidad de estudiantes representados en la muestra de 2015 hace que sus resultados no sean comparables.

4 <https://www.elpais.com.co/colombia/santos-destaca-mejora-en-pruebas-pisa-y-dice-que-debemos-sentirnos-orgullosos-de-resultados.html>

**Referencias Bibliográficas**

Alexander, R. (2001). Culture and Pedagogy. International Comparisons in Primary Education. Londres: Blackwell.

Ball, S.J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, no. 18, 215-228, DOI: 10.1080/0268093022000043065Berliner, D. (2011). Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, 41, 287-302, DOI:10.1080/0305764X.2011.607151.

Bos, M.S., Elías, A., Vegas, E. y Zoido, P. (2016). PISA: América Latina y el Caribe. ¿Cuánto mejoró la región? [Nota 2, Banco Interamicano de Desarrollo, Centro de Información para la mejora de los aprendizajes]. Recuperado el 10/10/2018, de <https://publications.iadb.org/bitstream/handle/11319/7992/America-Latina-en-PISA-2015-Cuanto-mejoro-la-region.pdf?sequence=4>

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

Carnoy, M., Khavenson, T., Fonseca, I., Costa, L. y Marotta, L. (2015). Is Brazilian education improving? Evidence from PISA and SAEB. *Cadernos de Pesquisa*, 45(157), 450-485, DOI:10.1590/198053143331.

Carnoy, M. y Rothstein, R. (2013). International tests show achievement gaps in all countries, with big gains for U.S. disadvantaged students. Washington DC: Economic Policy Institute.

Chen, Y.F., y Jiao, H. (2014). Exploring the utility of background and cognitive variables in explaining latent differential item functioning: An example of the PISA 2009 Reading assessment. *Educational Assessment*, 19(2), 77–96, DOI:10.1080/10627197.2014.903650.

Cosgrove, J. (2015). Changes in Achievement in PISA from 2000 to 2009 in Ireland: beyond the test scores. *The Irish Journal of Education*, xl, pp. 29-44, DOI:10.1186/2196-0739-2-2.

Cowen, R. (2011). “CODA”. En Pereyra, Kotthoff y Cowen (eds.). PISA Under Examination: Changing Knowledge, Changing Tests, and Changing Schools (pp. 259-264). Holanda: Sense Publishers.

Eivers, E. (2010). PISA: Issues in implementation and interpretation. *Irish Journal of Education/Iris Eireannach an Oideachais*, 38, 94–118. Recuperado el 15/10/2018, de <http://www.jstor.org/stable/20789130>

Eklöf, H. (2010). Skill and will: test‐taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, 17:4, 345-356, DOI:10.1080/0969594X.2010.516569.

Fernandez Cano, A. (2016). A Methodological Critique of the PISA Evaluations. *Relieve*, 22(1). DOI:10.7203/relieve.22.1.8806.

Ferrer, G. (2006). Sistemas de evaluación de aprendizajes en América Latina. Balances y desafíos. Santiago de Chile: Educar Chile.

Ferrer, G. y Fiszbein, A. (2015). “¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América Latina?”. Documento de Antecedentes. Comisión para la Educación de Calidad para Todos. El Diálogo Interamericano. Washington DC: World Bank Group.

Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, 1–17, DOI: 10.1002/ets2.12067

Fischman, G.E., Marcetti Topper, A., Silova, I., Goebel, J. y Holloway, J.L. (2018). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, 34:4, 470-499, DOI: 10.1080/02680939.2018.1460493.

Freitas, P., Catela Nunes, L., Balcão Reis, A., Seabra, C. y Ferro, A. (2016). Correcting for sample problems in PISA and the improvement in Portuguese students’ performance. *Assessment in Education: Principles, Policy & Practice*, 23:4, 456-472. DOI: 10.1080/0969594X.2015.1105784.

Fuller, B., Henne, M.K. y Hannum, E. (2008). Strong States, Weak Schools: The Benefits and Dilemmas of Centralized Accountability. UK: Emerald Group Publishing Limited.

Huerta Cuervo, M. Rocío (2010). Iberoamérica en PISA 2006: Informe regional del Grupo Iberoamericano de PISA. Gestión y Política Pública, 19(2), 420-430. Recuperado el 10/10/2018, de <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-10792010000200008>

Gneezy, U., List, J.A., Livingston, J.A., Sadoff, S., Qin, X. y Xu, Y. (2017). Measuring Success in Education: The Role of Effort on the Test Itself. *NBER Working Paper No*. 24004. DOI: 10.3386/w24004.

Grek, S., Lawn, M. y Ozga, J. (2009). Fabrication, circulation and use of a supra-national instrument of regulation based on knowledge –Education Sector: Production of OECD’s Programme for International Student Assessment (PISA). Edinburgh: Centre for Educational Sociology, University of Edinburgh.

Hanushek, E. y Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics*, 99(2), 497-512. DOI:10.1016/j.jdeveco.2012.06.004.

Hopfenbeck, T.N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J. y Baird, J. (2017). Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, 62 (3), 333–353. DOI:10.1080 /00313831.2016.1258726.

Instituto Colombiano para la Evaluación de la Educación (2017). Informe Nacional de Resultados. Colombia en PISA 2015. Ministerio de Educación de Colombia.

Instituto Nacional de Evaluación Educativa (2016). México en PISA 2015. México: INEE. Recuperado el 10/10/2018, de <http://publicaciones.inee.edu.mx/buscadorPub/P1/D/316/P1D316.pdf>

Jodouin, H. (2014). *Differential omission rates: Exploring factors that influence omission rates in a large-scale Pan-Canadian assessment*, PhD Thesis, Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto.

Klein, R. (2011). Uma re-análise dos resultados do PISA: problemas de comparabilidade*. Ensaio: Avaliação e Políticas Públicas em Educação*, 19 (73), 717-742, DOI:10.1590/S0104-40362011000500002

Kreiner, S. y Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, 79(2), 210-231, DOI :10.1007/s11336-013-9347-z.

Lewis, S. y Lingard, B. (2015). The Multiple Effects of International Large-Scale Assessment on Education Policy and Research. *Discourse*, 36 (5): 621–637. DOI:10.1080/01596306.2015.1039765.

Lindblad, S., Pettersson, D., y Popkewitz, T. S. (2015). International comparisons of school results: A systematic review of research on large-scale assessments in education. Stockholm: Swedish Research Council.

LLECE-UNESCO (2014). “Primera entrega de resultados”. TERCE: Tercer estudio regional, comparativo y explicativo. Santiago de Chile: OREALC-UNESCO.

Lüdtke, O., Robitzsch, A., Trautwein, U., Kreuter, F., e Ihme Marten, J. (2007). Are there test administrator effects in large-scale educational assessments? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 3(4), 149–159, DOI:10.1027/1614-2241.3.4.149

Meyer, H.-D., y Benavot, A. (2013). *PISA, power, and policy: The emergence of global educational governance*. UK: Symposium Books.

Ministério da Educação do Brasil (2016). Brasil no PISA 2015: análises e reflexões sobre o desempenho dos estudantes brasileiros. São Paulo: Fundación Santillana.

Mislevy, R. J. (1995). Test theory and language-learning assessment. *Language Testing*, 12(3), 341–369. DOI:10.1177/026553229501 200305.

Morris, A. (2011). Student Standardised Testing: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers, No. 65, OECD Publishing, Paris. Recuperado el 15/10/2018, de <http://dx.doi.org/10.1787/5kg3rp9qbnr6-en>

Niemann, D., Martens, K. y Teltemann, J. (2017). PISA and its Consequences: Shaping Education Policies through International Comparisons. *European Journal of Education*, 52 (2): 175–183, DOI:10.1111/ejed.12220

OECD (2015). PISA 2015: Techinal Standards. París: OCDE Publishing.

OECD (2016a). PISA 2015 Results (Volume I): Excellence and Equity in Education. París: OCDE Publishing. Recuperado el 15/10/2018, de <http://dx.doi.org/10.1787/9789264266490-en>.

OECD (2016b). Programme for International Student Assessment. Results from PISA 2015. Country Note: Colombia. OCDE Publishing. Recuperado el 15/10/2018 de <https://www.oecd.org/pisa/PISA-2015-Colombia.pdf>

OECD (2018). Competencias en Iberoamérica: análisis de PISA 2015. OCDE Publishing & Fundación Santillana –for spanish edition. Recuperado el 15/10/2018, de<https://www.segib.org/wp-content/uploads/Competencias-en-Iberoamerica-Analisis-de-PISA-2015.pdf>

Pintrich, P. R. & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. DOI:10.1037//0022–0663.82.1.33.

Pons, X. (2017). Fifteen Years of Research on PISA Effects on Education Governance: A Critical Review. *European Journal of Education*, 52 (2): 131–144, DOI:10.1111/ejed.12213

Rautalin, M., Alasuutari, P. y Vento, E. (2018). Globalisation of education policies: does PISA have an effect? *Journal of Education Policy*. 34:4, 500-522, DOI: 10.1080/02680939.2018.1462890

Ravela, P. (2003). ¿Cómo aparecen los resultados de las evaluaciones educativas en la prensa? Washington DC: PREAL.

Rivas, A. (2015). América Latina después de PISA: Lecciones aprendidas de siete países. Buenos Aires: CIPPEC.

Solano-Flores, G. y Milbourn, T. (2016). Capacidad evaluativa, validez cultural y validez consecuencial en PISA. *Relieve*, 22(1), DOI:10.7203/relieve.22.1.8281.

Swaffield, S y Thomas, S. (2016). Educational Assessment in Latin America. *Assessment in Education: Principles, Policy & Practice*, 23:1, 1-7, DOI: 10.1080/0969594X.2016.1119519.

Taubman, P. M. (2009). Deconstructing the Discourse of Standards and Accountability in Education. New York and London: *Routledge, Taylor and Francis Group*.

Wise, S.L., y DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and possible solutions. *Educational Assessment* 10(1), 1–17, DOI:10.1207/s15326977ea1001\_1.

Wolff, L. (2004). Educational Assessments in Latin America: the state of art. *International Association of Applied Psychology*, 53(2), 192-214, DOI:10.1111/j.1464-0597.2004.00168.x. 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. Recuperado el 15/10/2018, de <https://ssrn.com/abstract=1159042>

Wolff, L. y Smith, J.K. (1995). The consequence of consequence: motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227-242.

**Abstract**

**The Second Dimension of Educational Quality: A Comparative Study in Seven Countries in Latin America**

INTRODUCTION. The educational quality assessments made it possible to compare education systems in the last 20 years. PISA test results are the most commonly used as reference. In Latin America, there are seven countries that have recurrently participated in this test since the year 2000: Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Uruguay.

METHODOLOGY. In this study, we analyze the evolution of the above-named countries in the PISA test, and other national and international assessments. We examine the goals, methodological biases, and the comparative results of these tests.

RESULTS. The results of the educational quality assessments contradict each other and have variations based on methodological issues that are relatively unexplored. The PISA tests have introduced a methodological change in its 2015 edition that modify the results of previous editions and made the results incomparable to those of the first editions (2000 and 2003). This raises important questions regarding the fact that the educational systems that seem to have improved, have actually not, or at least not to such extent. The paper introduces a comparison of the results of different assessments in which those countries participated, showing some cases that have improved their educational quality like Peru—which had a lower starting point— and Chile— which has shown the highest improvement among the analyzed countries during the 2000s.

DISCUSSION. The study concludes by revealing a necessity to discuss a “second dimension of the educational quality” that requires further analyses of the standardized assessment results—not just skim reading— taking into consideration the various educational indicators and sources in a complimentary manner. That second dimension acts as an epistemological filter to understand the evolution of educational systems in Latin America and to avoid a misreading or a biased interpretation of the improvements achieved in the quality of education.

**Keywords**: Educational Testing, Educational Policy, Educational Quality, Comparative Analysis, re-scaling

**Résumé**

**La deuxième dimension de la qualité de l'éducation: Une étude comparée dans sept pays d'Amérique Latine**

INTRODUCTION. Les évaluations de la qualité de l’enseignement ont permis d’étudier au cours des 20 dernières années la comparaison des systèmes éducatifs. Les résultats des tests PISA sont la référence la plus utilisée. En Amérique Latine, sept pays y participent depuis l’an 2000: l’Argentine, le Brésil, le Chili, la Colombie, le Mexique, le Pérou et l’Uruguay.

MÉTHODE. On analyse ici l’évolution de ces pays dans le test PISA et dans d’autres évaluations nationales et internationales. Aussi les objectifs, les biais méthodologiques et les résultats comparatifs de ces évaluations sont analysés.

RÉSULTATS. Les résultats des différentes évaluations de la qualité font preuve des tendances qui montrent des contradictions entre elles, même avec des variations qui reposent sur des problèmes méthodologiques peu explorés jusqu'à présent. Les tests PISA ont introduit des modifications dans l'édition 2015 qui modifient les résultats des éditions précédentes et rendent les résultats des premières éditions (2000 et 2003) incomparables. On s'interroge sur le constat que les systèmes éducatifs qui semblaient s'être améliorés ne l'ont pas réellement fait ou l'ont réalisé de manière plus réduite qu'on ne le pensait auparavant. Une vue comparative des résultats des évaluations des sept pays analysés est présentée, confirmant qu'il y avait des cas d'amélioration tels que celui du Pérou à partir d'un point de départ inférieur et celui du Chili dans les années 2000, atteignant le point culminant parmi les pays analysés.

DISCUSSION. L'étude présente la nécessité de discuter une « deuxième dimension de la qualité de l'éducation », qui nécessite d'aller au-delà de la lecture simplifiée des résultats des évaluations, en considérant différentes sources et différents indicateurs de l'éducation de manière complémentaire. Cette deuxième dimension fonctionne comme un filtre épistémologique pour comprendre l'évolution des systèmes éducatifs en Amérique Latine et pour éviter les lectures erronées ou biaisées de la qualité de l'éducation.

**Mots-clés :** Évaluation de l'éducation, Politique de l'éducation, Qualité de l'éducation, Analyse comparative, Redimensionnement.

Axel Rivas (autor de contacto)

Profesor, Investigador y Director de la Escuela de Educación de la Universidad de San Andrés. Doctor en Ciencias Sociales, Universidad de Buenos Aires (UBA). Master en Ciencias Sociales y Educación, FLACSO, Argentina. Licenciado en Ciencias de la Comunicación (UBA). Autor de libros y artículos sobre perspectivas comparadas y políticas de la educación; mejora de los sistemas educativos, e innovación educativa.

e-mail: arivas@udesa.edu.ar

Dirección: Vito Dumas 284 (B1644BID), Victoria, Buenos Aires, Argentina.Teléfono: +5491131217072

e-mail: arivas@udesa.edu.ar

Martín Guillermo Scasso.

Licenciado y profesor en Sociología, Universidad de Buenos Aires. Consultor para la producción y análisis de información cuantitativa en el sector educación. Consultor de organismos de cooperación internacional (UNICEF, UNESCO, Banco Mundial, CECC/SICA, OEI). Fue asesor de la Secretaría de Evaluación Educativa del Ministerio de Educación y Deporte de Argentina.

e-mail: martinscasso@gmail.com