

## **SEXTO INFORME ESTADO DE LA EDUCACION**

### **¿Es “real” el descenso en los puntajes de Costa Rica en las pruebas PISA 2015?**

Eiliana Montero  
Estado de la Educación-PEN  
Universidad de Costa Rica

Enero, 2017



Nota: Las cifras de las ponencias pueden no coincidir con las consignadas por el Sexto Informe Estado de la Educación (2017) en el tema respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

## **1. Antecedentes y justificación**

Una de las metas principales para la mayoría de las naciones que participan en las pruebas del Programa Internacional para la Evaluación de Estudiantes (PISA, por su sigla en inglés) de la OCDE, más allá de conocer su posición en el *ranking* de los países involucrados, es dar seguimiento al desempeño de sus estudiantes en el tiempo, con el propósito de monitorear avances o retrocesos. Si otros factores del contexto permanecen constantes, y los resultados obtenidos muestran una tendencia creciente, la conclusión es que el sistema educativo está haciendo las cosas mejor que antes. Por el contrario, si el rendimiento se mantiene o disminuye, es señal de que debe hacerse un mayor esfuerzo en este ámbito.

Las pruebas PISA se aplican a estudiantes de 15 años y se enfocan en tres áreas principales: alfabetización matemática, competencia lectora y alfabetización científica. Se espera que, comparando directamente los puntajes obtenidos entre un año y otro, se pueda determinar si el desempeño de las y los alumnos en esas materias está mejorando o no. Sin embargo, hay condiciones metodológicas fundamentales que deben cumplirse para que estas comparaciones y conclusiones sean válidas.

En el caso de la prueba aplicada en Costa Rica en 2015, hay razones de peso para dudar de la validez de las comparaciones entre ese año y los anteriores. En otras palabras, no se puede concluir sobre mejoras o desmejoras en el rendimiento estudiantil a partir de la comparación directa entre los puntajes de 2015 y los de 2012 y 2009. Ello se debe a dos circunstancias concretas: la cobertura de la población de 15 años, que aumentó de manera relevante con respecto a las pruebas anteriores, y el cambio en el modo de administración de los exámenes, que pasó de lápiz y papel en 2009 y 2012, a computadora en 2015.

Considerar la limitación metodológica que en términos de comparabilidad introducen estos dos factores, es crucial para asegurar la validez de las interpretaciones que deben alimentar la toma de decisiones educativas, pues no hacerlo podría llevar a conclusiones totalmente equivocadas, como indicar que el rendimiento de las y los jóvenes costarricenses está disminuyendo en las áreas evaluadas por PISA.

De igual forma, esta situación sirve para ejemplificar la necesaria vigilancia a nivel técnico-científico que debe mantener el Ministerio de Educación Pública (MEP) en torno a esta importante evaluación internacional, junto con otras instancias de investigación académicas nacionales interesadas en el mejoramiento de la educación en el país.

## **2. Objetivos de estudio**

- Indagar si la disminución en los puntajes de PISA 2015 en Costa Rica puede deberse a los cambios en la cobertura y el modo de administración de las pruebas y no a un menor rendimiento de los estudiantes
- Proponer un marco de referencia y procedimientos para explorar en qué medida cambios en las condiciones de aplicación de las pruebas PISA pueden invalidar las

interpretaciones en términos del rendimiento de los estudiantes, derivadas de las comparaciones interanuales de los puntajes reportados.

### 3. Método

Para realizar el estudio no hace falta equilibrar todas las variables comunes en PISA 2009, 2012 y 2015, pues precisamente el Consorcio encargado de la prueba se asegura de que las muestras sean equivalentes año a año (excepto cuando aumenta la cobertura, por supuesto) y de que la dificultad de las pruebas sea controlada *a priori*, en el ensamblaje, y *a posteriori* con métodos de equiparación de puntajes, para así poder comparar de manera directa y válida los puntajes de un año a otro y concluir en términos de la tendencia del constructo de interés, ya sea alfabetización matemática, alfabetización científica o competencia lectora.

Todos estos procedimientos se describen en los manuales técnicos producidos por los expertos de PISA y justifican la comparación "en bruto" de los puntajes año a año. Si las muestras no fueran equivalentes en habilidad o las pruebas no se equipararán por dificultad, las diferencias en los puntajes no reflejarían, necesariamente, aumentos o disminuciones en el constructo de interés.

En el presente caso, lo que se debe tratar de controlar, en la medida de lo posible, son los factores que cambiaron en el año 2015 para nuestro país, que lamentablemente son dos y están "confundidos" desde el punto de vista estadístico: la cobertura y el modo de administración de la prueba, de lápiz y papel a administración por computadora.

Ya en el informe de PISA 2009, primer año en que Costa Rica participó en este programa, se hizo notar que la validez de las comparaciones sustantivas entre países se veía amenazada debido a la falta de comparabilidad de los puntajes reportados, causada a su vez por las grandes diferencias de cobertura entre las naciones participantes. En aquel momento se indicó:

*Sólo un 53% de los jóvenes de 15 años del país están siendo representados por la muestra de PISA del año 2009, las exclusiones las componen tres grupos, a saber: el sorprendentemente alto porcentaje de estudiantes que poseen adecuaciones en los colegios, aquellos que están en educación abierta y los que no están activos en el sistema educativo. Este es uno de los porcentajes de representación más bajos de los países participantes de PISA en ese año. Para establecer una referencia comparativa, los porcentajes de representación para México y Chile son respectivamente 61% y 85%. Es de esperar que el grupo de jóvenes excluidos del marco muestral tenga un desempeño relativamente menor en estas pruebas, comparado con el grupo que sí se muestreó (Montero et al., 2013).*

En esa oportunidad se siguió la metodología propuesta en el informe en español de PISA 2009 (OCDE-Santillana, 2011) y se realizó un ejercicio, que llamamos simulación, para determinar cuáles habrían sido los puntajes de Costa Rica en las dos pruebas que se estudiaron en esa ocasión –competencia lectora y alfabetización matemática– si el país

hubiera tenido, respectivamente, los niveles de representación de México (61%) y de Chile (85%). Los resultados obtenidos se presentan en el cuadro 1.

En la comparación con México se determinó que en la prueba de competencia lectora las diferencias significativas se mantenían a favor de Costa Rica, es decir, es decir nuestro país se ubicaba por encima de México; mientras que en alfabetización matemática sucedía lo contrario, el promedio mexicano era significativamente más alto que el costarricense.

Por el contrario, en la comparación con Chile, si Costa Rica hubiera tenido los mismos niveles de cobertura que este país en PISA 2009, sus promedios de rendimiento en ambas pruebas continuarían ubicándose por debajo de los de Chile, pero con una diferencia estadística significativamente mayor.

**Cuadro 1**

**Costa Rica, Chile y México: resultados comparativos entre los promedios en las pruebas 2009 PISA simulando para Costa Rica la cobertura de Chile y México**

Caso	Competencia lectora		Alfabetización matemática	
	Promedio	Error estándar	Promedio	Error estándar
Chile	449	3,1	421	3,1
Costa Rica (a)	424	6	390	6
Costa Rica observado	443	3,2	409	3
México	425	2	419	1,8
Costa Rica (b)	438	3,2	404	3

(a) Simulando para Costa Rica la cobertura de Chile (85%)

(b) Simulando para Costa Rica la Cobertura de México (61%)

Fuente: Montero et al., 2012

Esta simulación evidenció que la comparación directa entre los puntajes promedio brutos de Costa Rica y Chile debe hacerse con cautela, pues las diferencias de cobertura hacen cuestionable la validez de ese ejercicio. Así, no es correcto afirmar que, en términos poblacionales, los desempeños de Costa Rica y Chile en PISA 2009 fueron iguales, ya que no existen diferencias estadísticamente significativas entre sus promedios reportados. Una última observación con respecto al ejercicio desarrollado años atrás, con los datos de PISA 2009, es que en 2015 Costa Rica tuvo una cobertura de 63%, similar al 61% de México en 2009, y sus resultados se acercaron bastante a los puntajes predichos por la simulación, pues los promedios fueron 427.05 y 397.31 para competencia lectora y alfabetización matemática, respectivamente.

En consecuencia, se aplicó la metodología a los datos de 2015, para estimar cuáles hubieran sido los puntajes en las tres materias si en 2009 y 2012 la cobertura hubiera sido la misma de 2015 (63%). Con esto estamos controlando el primer factor, o sea, el aumento

en la cobertura. Así, los puntajes promedio disminuyen 6 puntos en el 2009 y 8.04 puntos en el 2012.

Aun después de controlar este factor queda variancia irrelevante al constructo que debe controlarse y que invalida las comparaciones entre el 2015 y los años anteriores: la modificación en la forma de administrar la prueba. Que el cambio de papel y lápiz a la computadora reduzca los puntajes promedio tiene sentido en un contexto como el de Costa Rica, un país en vías de desarrollo, pues hay diferencias importantes en el grado de alfabetismo computacional de las y los estudiantes. A nivel internacional hay abundante literatura que documenta este fenómeno y, para nuestro caso, uno de los estudios más relevantes fue realizado por John Jerrim, del Instituto de Educación de la University College de Londres, con datos de PISA 2012. En este el autor indica (el subrayado es nuestro):

*Our results provide evidence of important and interesting differences. Despite a strong cross-country correlation, mean paper and computer test scores differ by at least 10 PISA points in one-in-three economies. There is a particularly notable decline in Shanghai-China, where children score 50 points lower (on average) on the computer-based mathematics test (Jerrim, 2016, p.6).*

Otra conclusión de este estudio, que parece muy apropiada en nuestro contexto, es que el efecto del modo de administración no es igual para todas las categorías de población, sino que hay ciertos grupos que se ven más afectados que otros.

Así pues, para llevar a cabo el análisis que aquí se presenta se adoptó un enfoque conservador y se supuso que la disminución en los puntajes reportados de PISA debido a la administración del examen por computadora es igual a diez puntos. Esto implica restar esa cantidad a los puntajes del 2009 y 2012 antes de hacer la comparación.

## 1. Resultados

Siguiendo esta aproximación, un primer análisis de variancia y comparaciones *post hoc* con el primer valor plausible para cada estudiante<sup>1</sup>, original (sin ajustar) para los años 2009, 2012 y 2015, muestra los resultados que se presentan en siguientes cuadros.

---

<sup>1</sup> Es importante tener presente que en el 2015, el Consorcio encargado de PISA estimó el desempeño de los alumnos calculando diez valores plausibles para cada estudiante en cada prueba. El valor plausible es la estimación del desempeño que hubiera tenido el estudiante si hubiera contestado todos los ítems de la prueba y no una sub-muestra como en la práctica ocurre; esto por cuanto existe más de una veintena de cuadernillos diferentes y no todos los alumnos contestan los mismos contenidos. Las correlaciones entre esos valores plausibles son muy altas, superiores a 0.85.

**Cuadro 2**

**Conjuntos homogéneos de medias. Primer Valor Plausible original, sin ajustar Costa Rica, aplicaciones 2009, 2012 y 2015, Prueba PISA de Lectura**

	Periodo	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	427.0524	
	2012	4391		440.7301
	2009	4578		443.3033
	Sig.		1.000	.072
Scheffe <sup>a,b</sup>	2015	6866	427.0524	
	2012	4391		440.7301
	2009	4578		443.3033
	Sig.		1.000	.197

**Cuadro 3**

**Conjuntos homogéneos de medias. Primer Valor Plausible original, sin ajustar Costa Rica, aplicaciones 2009, 2012 y 2015, Prueba PISA de Matemática**

	Periodo	N	Subset for alpha = 0.05		
			1	2	3
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	397.3101		
	2012	4391		405.8904	
	2009	4578			408.8731
	Sig.		1.000	1.000	1.000
Scheffe <sup>a,b</sup>	2015	6866	397.3101		
	2012	4391		405.8904	
	2009	4578		408.8731	
	Sig.		1.000	.063	

**Cuadro 4****Conjuntos homogéneos de medias. Primer Valor Plausible original, sin ajustar Costa Rica, aplicaciones 2009, 2012 y 2015, Prueba PISA de Ciencias**

	Periodo	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	414.8000	
	2009	4578		429.6277
	2012	4391		430.2747
	Sig.		1.000	.622
Scheffe <sup>a,b</sup>	2015	6866	414.8000	
	2009	4578		429.6277
	2012	4391		430.2747
	Sig.		1.000	.886

Como se puede notar, en comprensión lectora los datos sin ajustar revelan diferencias significativas entre los promedios de 2015 y los años anteriores. En alfabetización matemática el análisis arroja tres promedios distintos a nivel poblacional para los tres años, tendiendo hacia la baja, y en alfabetización científica el promedio del 2015 se diferencia significativamente y hacia abajo con respecto a los dos años anteriores.

Sin embargo, estas comparaciones no son válidas para formular conclusiones sobre el rendimiento de los estudiantes, porque en los tres años hubo coberturas distintas y, además, en 2015 también cambió el modo de administración de la prueba.

En vista de lo anterior, se hace una primera simulación siguiendo el procedimiento aplicado en el informe de 2012 y ajustando los promedios de los años 2009 y 2012, de manera que reflejen la misma cobertura de 2015 (63%). Esto genera una caída en los promedios originales en el 2009 y 2012, de 6 y 8.04 puntos respectivamente.

**Cuadro 5****Conjuntos homogéneos de medias. Primer Valor Plausible ajustado por cobertura Costa Rica, aplicaciones 2009, 2012 y 2015, Prueba PISA de Matemática**

PV1MATH

	Periodo	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	397.02424	
	2012	4391	397.99420	
	2009	4578		403.23389
	Sig.		.479	1.000
Scheffe <sup>a,b</sup>	2015	6866	397.02424	
	2012	4391	397.99420	
	2009	4578		403.23389
	Sig.		.779	1.000

**Cuadro 6**  
**Conjuntos homogéneos de medias. Primer Valor Plausible ajustado por cobertura Costa Rica, aplicaciones 2009, 2012 y 2015, Prueba PISA de Lectura**

PV1READ

	Periodo	N	Subset for alpha = 0.05		
			1	2	3
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	426.48867		
	2012	4391		432.68023	
	2009	4578			437.52184
	Sig.		1.000	1.000	1.000
Scheffe <sup>a,b</sup>	2015	6866	426.48867		
	2012	4391		432.68023	
	2009	4578			437.52184
	Sig.		1.000	1.000	1.000

**Cuadro 7**

**Conjuntos homogéneos de medias. Primer Valor Plausible ajustado por cobertura Costa Rica, aplicaciones 2009, 2012 y 2015. Prueba PISA de Ciencias**

PV1SCIE				
	Periodo	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls <sup>a,b</sup>	2015	6866	414.60372	
	2012	4391		422.73602
	2009	4578		423.38672
	Sig.		1.000	.647
Scheffe <sup>a,b</sup>	2015	6866	414.60372	
	2012	4391		422.73602
	2009	4578		423.38672
	Sig.		1.000	.900

En este contexto, concluimos que en el caso de alfabetización matemática no se puede rechazar la hipótesis nula de igualdad de promedios poblacionales entre 2012 y 2015, si bien la comparación de 2009 con 2012 y 2015 sí resulta significativa, a favor del primer año. Para competencia lectora el análisis arroja tres promedios poblacionales distintos, tendiendo hacia la baja, y para alfabetización científica solo el promedio del 2015 se aleja significativamente y hacia abajo con respecto a los otros dos.

No obstante, las comparaciones entre el 2015 y los otros años están afectadas por el modo de administración de la prueba, ya que las posibles deficiencias en el alfabetismo computacional de ciertos grupos pueden provocar una desmejora en los puntajes observados, que no obedece a cambios en los conocimientos y habilidades de los estudiantes en la materia de interés. El escenario donde ese efecto promedio es relativamente bajo (10 puntos en la escala de PISA) se presenta en los siguientes cuadros.

**Cuadro 8**

**Conjuntos homogéneos de medias Primer Valor Plausible ajustado por cobertura y por modo de administración del test Costa Rica, aplicaciones 2009, 2012 y 2015. Prueba PISA de Matemática**

PV1MATH\_S2

	Periodo	N	Subset for alpha = 0.05		
			1	2	3
Student-Newman-Keuls <sup>a,b</sup>	2012	4391	387.99420		
	2009	4578		393.23389	
	2015	6866			397.02424
	Sig.		1.000	1.000	1.000
Scheffe <sup>a,b</sup>	2012	4391	387.99420		
	2009	4578		393.23389	
	2015	6866			397.02424
	Sig.		1.000	1.000	1.000

Cuadro 9

Conjuntos homogéneos de medias. Primer Valor Plausible ajustado por cobertura y por modo de administración del test Costa Rica, aplicaciones 2009, 2012 y 2015. Prueba PISA de Competencia Lectora

PV1READ\_S2

	Periodo	N	Subset for alpha = 0.05	
			1	2
Student-Newman-Keuls <sup>a,b</sup>	2012	4391	422.68023	
	2015	6866		426.48867
	2009	4578		427.52184
	Sig.		1.000	.498
Scheffe <sup>a,b</sup>	2012	4391	422.68023	
	2015	6866		426.48867
	2009	4578		427.52184
	Sig.		1.000	.795

Cuadro 10. Conjuntos homogéneos de medias. Primer Valor Plausible ajustado por cobertura y por modo de administración del test. Costa Rica, aplicaciones 2009, 2012 y 2015. Prueba PISA de Ciencias

PV1SCIE\_S2

	Periodo	N	Subset for alpha = 0.05
			1
Student-Newman-Keuls <sup>a..b</sup>	2012	4391	412.7360
	2009	4578	413.3867
	2015	6866	414.6037
	Sig.		.386
Scheffe <sup>a..b</sup>	2012	4391	412.7360
	2009	4578	413.3867
	2015	6866	414.6037
	Sig.		.421

En este análisis, al contrario de lo que muestran los resultados sin ajustar, el puntaje de alfabetización matemática en 2015 es significativamente superior a las calificaciones de 2009 y 2012. En competencia lectora no hay diferencias de relevancia entre 2009 y 2015, mientras que la cifra de 2012 se ubica más abajo. Por último, en alfabetización científica no se puede rechazar la hipótesis nula de igualdad de promedios poblacionales para los tres años.

#### 4. Conclusiones

- Se concluye que, en efecto, hay evidencia para afirmar que el rendimiento de las y los estudiantes costarricenses en PISA 2015 no fue más bajo que el de los años anteriores. Los descensos observados en los puntajes pueden haber sido causados por el aumento de la cobertura en 2015 y el cambio en el modo de administrar las pruebas (de papel y lápiz a computadora).
- En el caso de alfabetización matemática incluso podría haberse registrado un leve incremento en el nivel de desempeño.
- El siguiente cuadro resume los resultados que dan origen a estas conclusiones sustantivas.

#### Cuadro 11

**Comparación de los puntajes promedio de Costa Rica en Pruebas PISA en Alfabetización Matemática, Competencia Lectora y Alfabetización Científica, según distintos escenarios, para hacer equivalente al 2015 las aplicaciones 2009, 2012 y 2015**

Año	Tipo de medida	Alfabetización Matemática	Competencia Lectora	Alfabetización Científica
2009	Promedio original sin ajustar	408.87	443.30	429.63

	Promedio ajustado por cobertura	403.23	437.52	423.39
	Promedio ajustado por cobertura y modo de administración	<u>393.23</u>	<u>427.52</u>	<u>413.38</u>
2012	Promedio original sin ajustar	405.89	440.73	430.28
	Promedio ajustado por cobertura	397.99	432.68	422.74
	Promedio ajustado por cobertura y modo de administración	<u>387.99</u>	<u>422.68</u>	<u>412.73</u>
2015	Promedio original sin ajustar	397.02	426.49	414.6
	Promedio ajustado por cobertura	397.02	426.49	414.6
	Promedio ajustado por cobertura y modo de administración	<u>397.02</u>	<u>426.49</u>	414.6
Pruebas de significancia estadística para los puntajes ajustados por las dos condiciones son a favor de ...		2015 con respecto a los otros 2 años, y 2009 con respecto a 2012	2015 y 2009 con respecto a 2012	No hay diferencias estadísticamente significativas

(\*) PISA estima el desempeño de los alumnos calculando diez valores plausibles para cada estudiante en cada prueba. El valor plausible es la estimación del desempeño que hubiera tenido el estudiante si hubiera contestado todos los ítems de la prueba y no una sub-muestra como en la práctica ocurre; por cuanto existe más de una veintena de cuadernillos diferentes y no todos los alumnos contestan los mismos contenidos. El análisis se hizo con los 10 valores pero el cuadro solo muestra el primero.

Fuente: Elaboración propia, 2017

## 5. Recomendación

Se recomienda a los responsables de las pruebas PISA en el MEP que valoren la adopción de este marco general de referencia y procedimientos para explorar, antes de la divulgación de los resultados, en qué medida cambios en las condiciones de aplicación de los exámenes pueden invalidar las conclusiones sobre el rendimiento de las y los estudiantes, derivadas de las comparaciones interanuales de los puntajes observados. Asimismo, se recomienda al MEP que informe al público sobre esta temática e indiquen, como autoridad competente, cuáles son las interpretaciones adecuadas que deben darse a los resultados, según las circunstancias específicas de cada año y su grado de comparabilidad con los anteriores.

## 6. Referencias bibliográficas

1. Jerrim, J. (2016). *PISA 2012: How do results for the paper and computer tests compare?* Working Paper No. 16-02. London, University College of London, Institute of Education.
2. Montero, E., Rojas, S., Rodino, A. y Zamora, E. (2013). *Costa Rica en las pruebas PISA 2009 de competencia lectora y alfabetización matemática*. Ponencia preparada para el *Cuarto Informe Estado de la Educación*. San José, Programa Estado de la Nación, Consejo Nacional de Rectores.
3. OCDE-Santillana (2011). *Informe PISA 2009. Lo que los estudiantes saben y pueden hacer. Rendimiento de los estudiantes en lectura, matemáticas y ciencias*. Vol. I. OCDE, Santillana y Ministerio de Educación de España. Edición en inglés, 2010; edición en español, 2011.