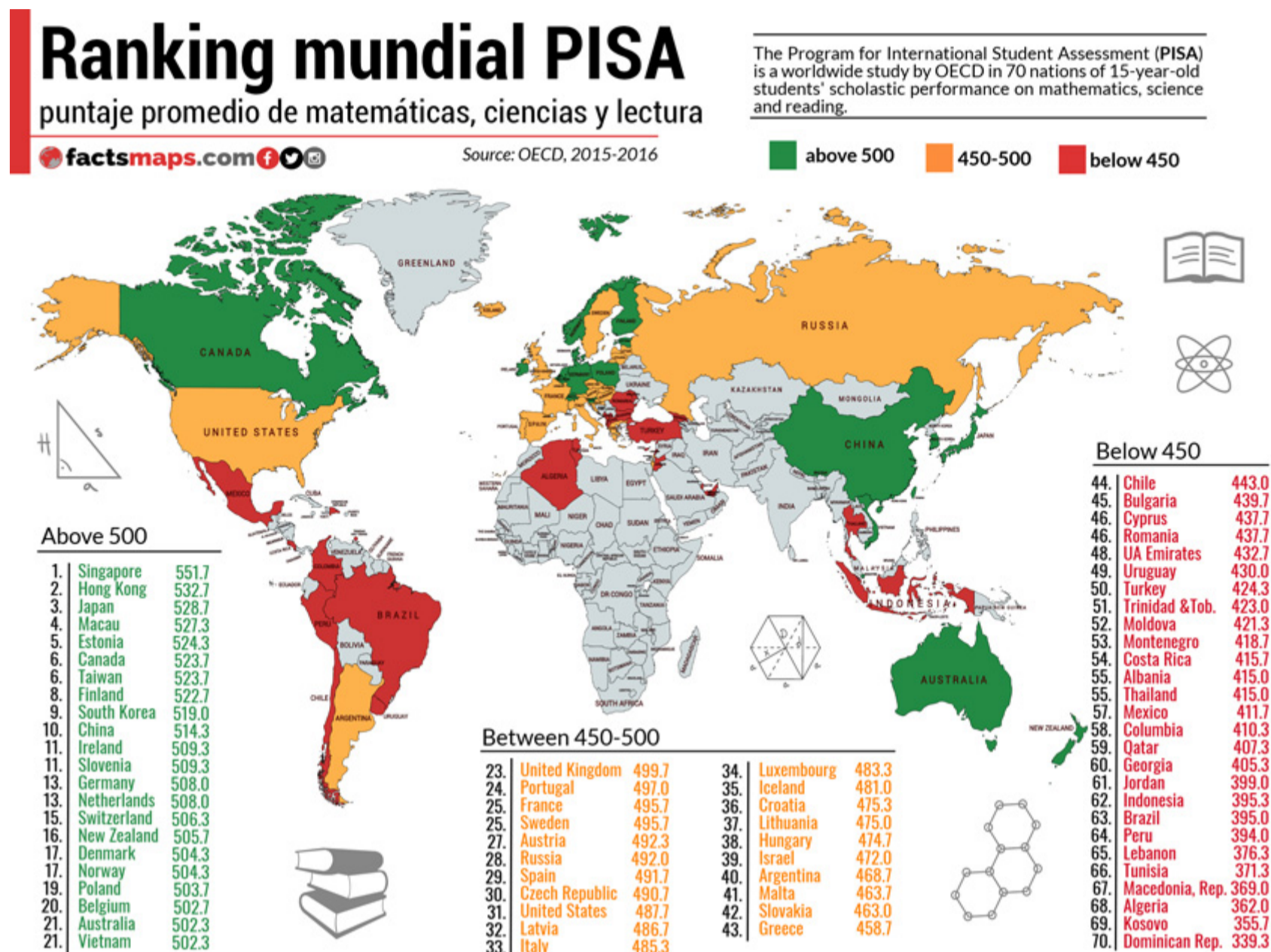


PISA 2018: one test to rule them all

4 de diciembre de 2019

Blog Satyagraha, 03 diciembre, 2019
Publicado por César Guadalupe

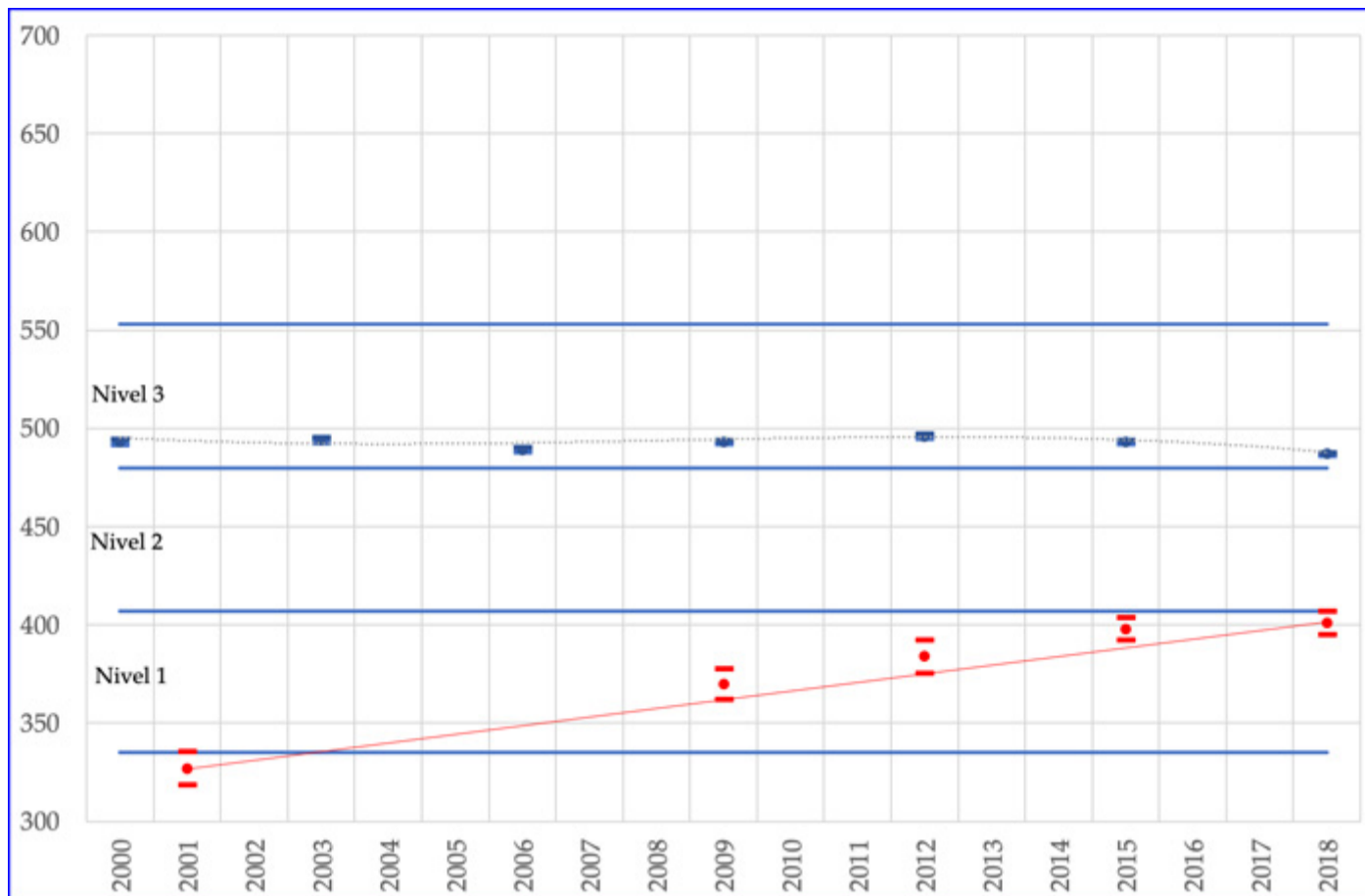
El día de hoy se publicaron los resultados de la ronda 2018 del Programa Internacional para la Evaluación de Estudiantes (PISA, sus siglas en inglés) del que participa el Perú (rondas de 2000, en 2001, y las de 2009, 2012, 2015 y 2018).



Para el caso del Perú, estos resultados básicamente muestran lo que se presenta a continuación; sin embargo, es preciso leer estos datos con mucha cautela. Por ello, luego de presentarlos, incluyo algunos elementos que me parece crucial tener en consideración.

Lectura: seguimos mejorando al mismo ritmo que hemos observado desde 2001. De hecho, parecería que entre 2015 y 2018 habría un estancamiento pero dados los atributos de PISA (que sólo se enfocó en lectura en 2000/2001, 2009, y 2018 los datos de 2015 no son muy sólidos a lo que habría que agregar la intervención hecha por el Ministerio que podría haber «inflado» esos resultados un poquito). Así, si en 2001 nos encontrábamos dos desviaciones estándar por debajo de la media de la OECD, ahora hemos reducido esa distancia a la mitad. Es decir, si seguimos a ese ritmo (que es tímido y, sin embargo, somos de los pocos países que mostramos algo así y, sobretodo de una forma sostenida a lo largo del tiempo) alcanzaríamos la media de la OECD en unos 20 años más. Nótese, que la media de la OECD está en caída (leve, pero caída al fin). Asimismo, si antes la media del Perú se encontraba debajo del nivel de desempeño más bajo que se había identificado entonces, ahora empezamos a llegar al segundo nivel de desempeño que podría ser considerado un mínimo aceptable (al decir de la propia OECD en este informe).

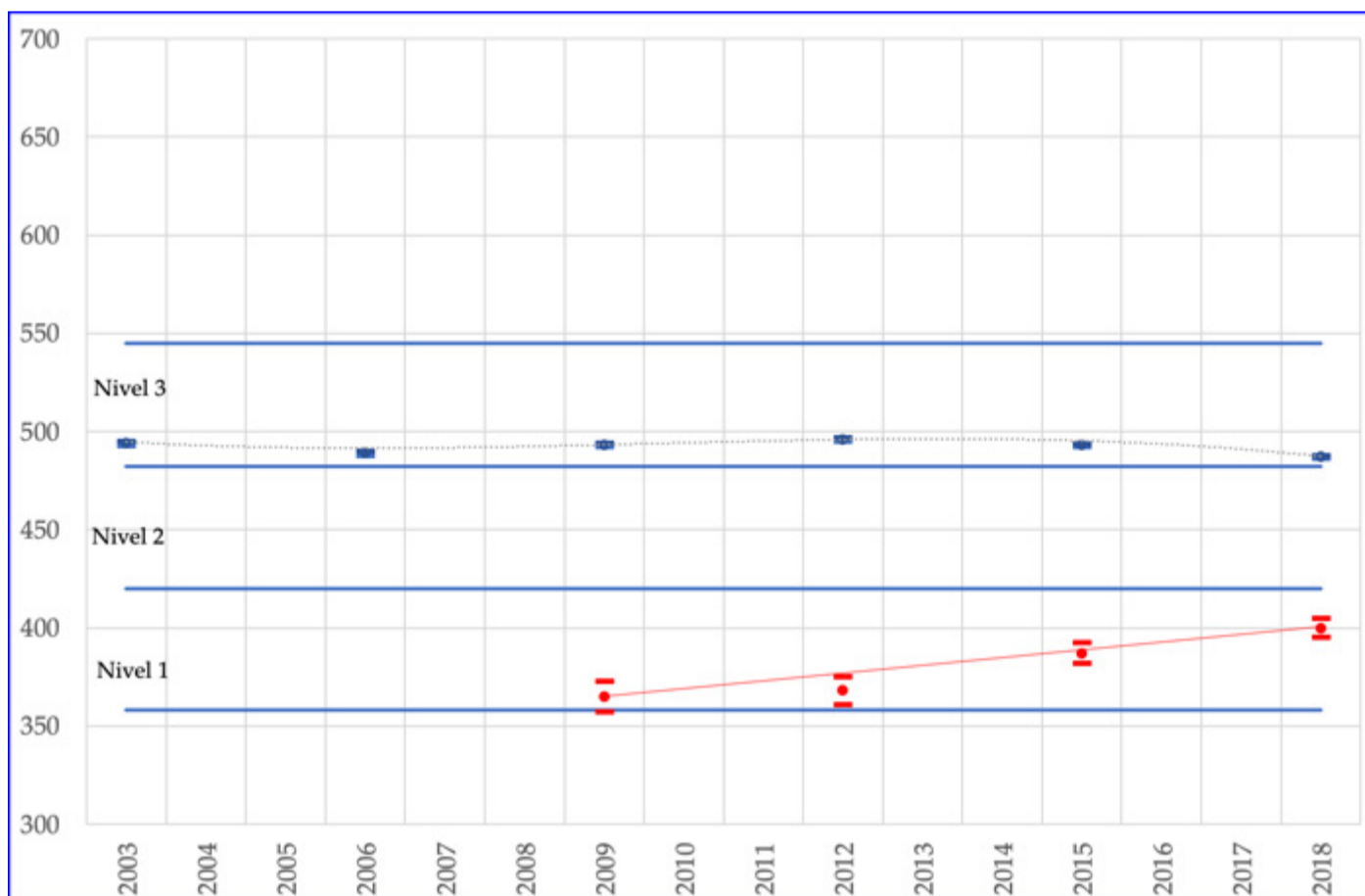
Estos resultados pueden observarse en el siguiente gráfico:



Puntajes medios de la OECD y del Perú en la escala combinada de lectura de PISA. 2000-2018. Las líneas horizontales marcan los límites superiores de los primeros tres niveles de desempeño definidos en 2000. Los puntos marcan el valor del estimador y las líneas los límites de los intervalos de confianza de la estimación (alpha: 0,05).

Matemáticas: Al igual que en lectura, vemos un progreso sostenido en el tiempo (aunque aquí sólo se puede tomar los datos peruanos desde 2009 y los de OECD desde 2003). Si en 2009 estábamos a poco menos de 1,5 desviaciones estándar de la media de la OECD, ahora estamos a una desviación estándar de distancia (con un promedio de la OECD también a la baja). Sin embargo, en este caso seguimos debajo del segundo nivel de desempeño que ahora la OECD considera como mínimo aceptable.

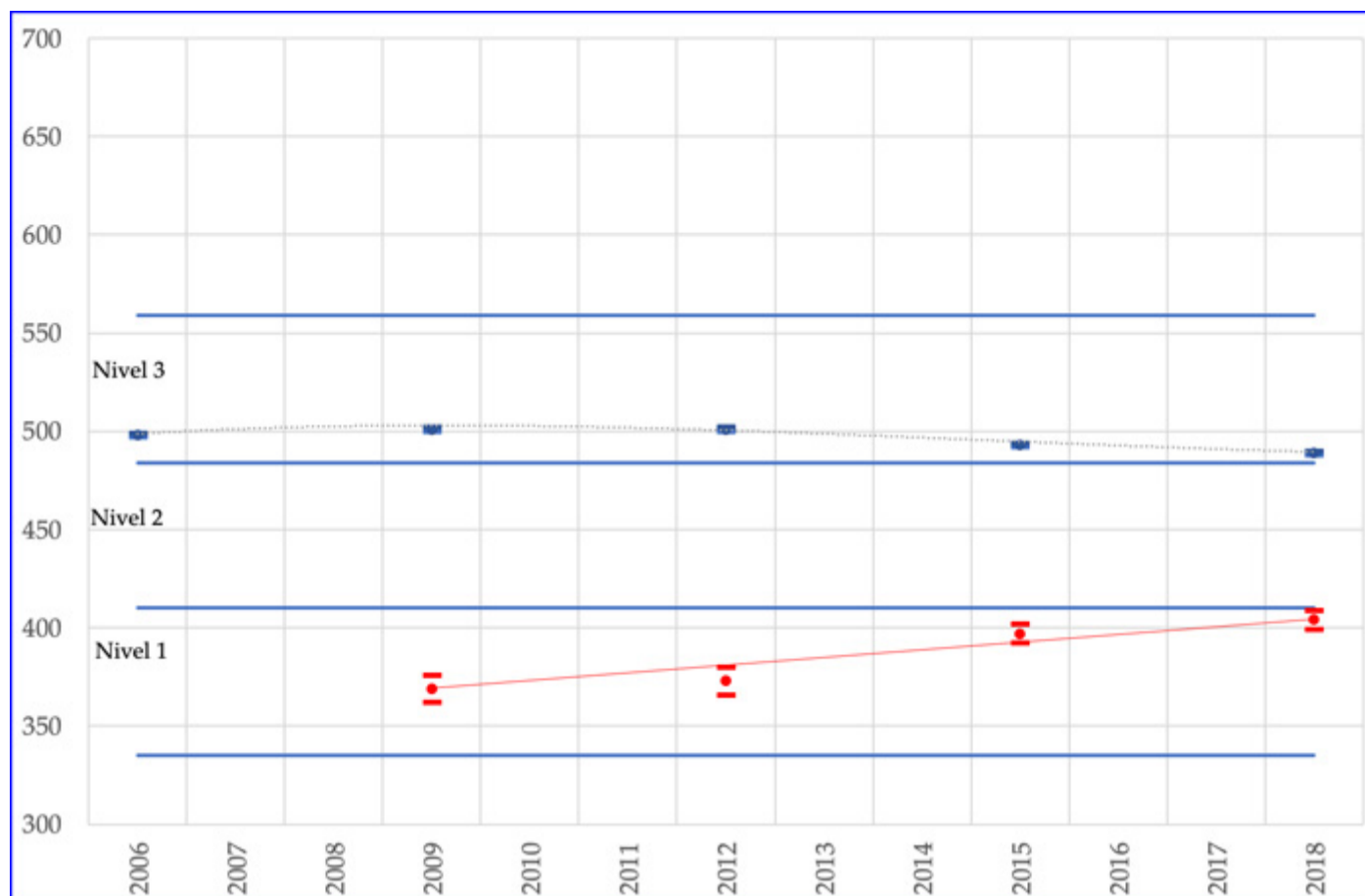
Estos resultados pueden observarse en el siguiente gráfico:



Puntajes medios de la OECD y del Perú en la escala combinada de matemáticas de PISA. 2000-2018. Las líneas horizontales marcan los límites superiores de los primeros tres niveles de desempeño definidos en 2000. Los puntos marcan el valor del estimador y las líneas los límites de los intervalos de confianza de la estimación (alpha: 0,05).

Ciencias: Al igual que en matemáticas, vemos un progreso sostenido en el tiempo (sólo se puede tomar los datos peruanos desde 2009 y de la OECD desde 2006). Si en 2009 estábamos a poco menos de 1,5 desviaciones estándar de la media de la OECD, ahora estamos a una desviación estándar de distancia (con un promedio de la OECD también a la baja). En este caso, como en lectura, nuestra media empieza a alcanzar el segundo nivel de desempeño que ahora la OECD considera como mínimo aceptable.

Estos resultados pueden observarse en el siguiente gráfico:



Puntajes medios de la OECD y del Perú en la escala combinada de ciencias de PISA. 2000-2018. Las líneas horizontales marcan los límites superiores de los primeros tres niveles de desempeño definidos en 2000. Los puntos marcan el valor del estimador y las líneas los límites de los intervalos de confianza de la estimación (alpha: 0,05).

Equidad: Nuevamente, el Perú aparece como uno de los países con mayores niveles de segregación escolar (las escuelas son homogéneas en términos socioeconómicos lo que impide fomentar el encuentro entre diversos y, con ello, mayor cohesión social) y donde la desventaja económica va unida a un menor desempeño en la prueba (es el país participante en PISA donde la condición socio-económica explica con mayor intensidad las variaciones en resultados). Es decir, tenemos escuelas para pobres en la que sólo hay personas pobres que logran pobres resultados. Nuestro sistema educativo no funciona para crear oportunidades para todos y asegurar algún nivel de movilidad social, más bien, reproduce las desigualdades.

Jóvenes y adultos: una novedad de este informe es que ahora se cuenta con información del Programa Internacional para la Evaluación de Competencias de los Adultos (PIAAC en inglés). Así, la OECD presenta información que relaciona los resultados de los jóvenes de 15 años, con los resultados de la generación de 35 a 55 años en el mismo país (presumiblemente, sus padres). Como cabe esperar, los resultados de los padres se asocian al resultado de los hijos. Lo triste en el caso peruano (datos de 2017 cuando se aplicó la prueba) es que para este grupo de personas nos encontramos muy lejos del promedio de la OECD. Ciertamente, esto se puede ver con mayor detalle en el informe correspondiente publicado el mes pasado (disponible en inglés aquí), y lamentablemente corrobora lo que anticipamos con dos colegas en un breve documento que publicamos hace un año (La larga noche de la educación peruana) y muestra como los problemas educativos de un momento dado se proyectan en el tiempo de una manera muy compleja. No hay forma de subestimar la importancia que tiene para el país que la mayoría de nuestra población adulta haya tenido una experiencia educativa tan pobre y presente un nivel tan limitado de habilidades básicas.

Toda esta información debe leerse tomando en cuenta lo siguiente:

PISA se inició como una prueba que quería medir las habilidades básicas de aquéllos que estaban a punto de ingresar a la fuerza laboral (por eso, personas de 15 años) independientemente de su nivel educativo. Así, PISA debía complementar la información sobre habilidades de la fuerza laboral (población de 16 a 65 años) que en los años 90 generó la Encuesta Internacional de Alfabetismo Adulto (IALS en inglés, hoy llamada PIAAC).

Dado el nivel de cobertura de los sistemas educativos de la OECD, se decidió que la mejor forma de aproximarse a la población de 15 años (la más eficiente) era ir a las instituciones educativas. Esto funcionaba bien para la mayor parte de los países de la OECD (excepto, entonces, México) pero se traduce en una gran dificultad para decir algo sobre los sistemas educativos ya que no se toma en cuenta las diferentes trayectorias de los estudiantes. Por ejemplo, en sistemas con altos niveles de atraso escolar (como el nuestro) la muestra toma estudiantes matriculados en todos los grados de la educación secundaria y que, por lo tanto, han tenido muy diversas oportunidades de desarrollar las habilidades medidas por la prueba. Así, tan pronto PISA inició sus operaciones, muchas personas observamos que tenía profundos problemas y limitaciones de cara a decir algo relevante sobre los sistemas educativos.

Sin embargo, el tremendo éxito comunicacional de PISA tornó a esta prueba en la estrella educativa de la OECD y en un "producto" que les permitía posicionarse como los que fijaban el estándar internacional en estos menesteres. Esto era particularmente importante (para la OECD) en un contexto en el que la Asociación Internacional para la Evaluación del Logro Estudiantil (IEA en inglés), que tenía una basta experiencia en conducir estudios de este tipo (desde su fundación en 1958), llevaba a cabo estudios que no obedecen a las preocupaciones fundamentalmente económicas de la OECD, sino educativas; sin embargo, los estudios de la IEA (que tienen base curricular y se enfocan en puntos específicos de dichas trayectorias por lo que son más útiles para desarrollar políticas educativas) nunca han logrado esa presencia comunicacional por tratarse de un consorcio de investigación y no de un organismo político con mayores recursos (con seguridad la mayoría de las personas que lean este post han oído hablar de PISA y la OECD, pero no de la IEA y estudios como PIRLS, TIMSS, ICCS, etc.). Es en ese marco, en el que la OECD no sólo lanza su estudio, sino que busca deslegitimar las alternativas presentando PISA como la única forma de lograr una adecuada compatibilidad internacional. Esto, por supuesto, es absurdo ya que la validez de un enfoque se mide contra los propósitos que busca servir y dado que los propósitos son diferentes se necesita enfoques diferentes.

En esta nueva ronda de PISA no sólo se insiste terca y absurdamente en ello, sino que también se fuerzan textos para afirmar que PISA es la mejor manera de medir uno de los indicadores de los Objetivos de Desarrollo Sostenible de las Naciones Unidas. Nótese que los textos que se fuerzan, claramente hablan de la necesidad de incluir varias aproximaciones y, en general, favorecen la medición de aprendizajes en puntos específicos de la trayectoria escolar, no de acuerdo a edades (véase el indicador 4.1.1 en la p. 24 de este documento).

Esta forma de proceder da cuenta de un propósito político detrás de PISA y que, crecientemente, ha adquirido más importancia dejando de lado la

propia robustez del estudio. La búsqueda de hegemonía e influencia es, hoy, más importante que la rigurosidad en PISA y eso debe quedarnos muy claro. PISA no es un estudio, es un instrumento político.

Si bien la información dada por PISA es muy limitada para definir algo que sea relevante para las políticas educativas (por su mirada cerrada sobre algunas habilidades, por no tomar en cuenta trayectorias educativas, etc.), era útil con esas limitaciones. Su utilidad consistía básicamente en lo siguiente: (i) nos permitía compararnos con los países de la OECD, (ii) nos permitía ver como esa comparación varía en el tiempo (si nos acercamos o nos alejamos), y (iii) permitía ver disparidades en el desempeño de los estudiantes dentro del país y constatar si somos más o menos equitativos que otros. Sin embargo, desde 2015 se han sumado algunos problemas adicionales que comprometen esa utilidad.

Listemos algunos de esos problemas:

1. desde 2015 el tratamiento de la no respuesta ha cambiado (antes se consideraba que una pregunta no respondida era errada, ahora se trata como omitida o no aplicada), esto es muy importante para países como el nuestro;
2. los procedimientos sicométricos usados (el modelo, la manera como se calibra la dificultad de las preguntas, la manera como se identifica problemas de comportamiento diferencial de los dichas preguntas) han variado;

Estos cambios, como ha sido mostrados por Rivas y Scasso (ver documento aquí) han afectado de modo marcado a los países latinoamericanos y fuerza a reprocesar toda la información anterior a 2015 para que sea realmente comparable. Lamentablemente, por disponibilidad de información los autores sólo han trabajado con información de 2006 en adelante. Siguiendo sus resultados, es posible afirmar con total certeza que si los datos de Perú en 2001 se hubiesen procesado de acuerdo a los cambios de 2015, no hubiesen sido tan malos y, por lo tanto, el ritmo de mejora observado sería más tímido que el que se muestra en los gráficos antes presentados.

Por si esto fuera poco, en 2018 se introdujeron algunos cambios adicionales muy importantes:

1. se ha cambiado la definición de «lectura» de modo que ahora se mide algo distinto que antes; además, se ha incluido en la medición algo que no puede ser usado para comparaciones internacionales: una medida de fluidez, que es dependiente de las características del idioma (por ejemplo, si la oración se estructura siguiendo la pauta Sujeto-Verbo-Objeto o si el verbo va al final; si hay mayor o menor grado de transparencia ortográfica; si hay más palabras monosilábicas -debido a que la lengua es tonal- y la forma como se aglomeran palabras, etc.)
2. desde 2018 se usa una administración adaptativa de la prueba basada en computadoras habiéndose abandonado la aplicación en papel.

Es decir, en un sentido estricto de las cosas, los datos de PISA de 2000 a 2012 son una cosa, los de 2015 otra, y los de 2018 una tercera. Es decir, la serie histórica de datos es, básicamente, inutilizable.

A lo anterior se suma que la alta visibilidad e impacto mediático de PISA ha llevado a varios países a malas prácticas. Por ejemplo, Colombia tiene un programa de entrenamiento para la prueba (ver acá), Escocia hace campañas de motivación para el desempeño (ver acá), el Ministerio peruano intervino (al parecer no con mucho éxito) en la muestra de 2015 afectando al motivación y la preparación de los estudiantes; entre otros ejemplos que podrían incluir las “anomalías” que llevaron a la OECD a no publicar los resultados españoles en lectura este año. En todos estos casos, se corre el riesgo que la muestra deje de ser lo que debe ser: una buena representación de la población bajo estudio con lo que los datos que arroja el estudio terminan siendo, simplemente, inservibles.

Finalmente, ciertas visiones de la Educación sumadas a los bajos niveles de competencias para leer este tipo de informes han llevado a mucha gente a priorizar la observación de los rankings de países (que es lo primero que presenta la propia OECD) como si eso diera alguna información relevante (si el conjunto de participantes varía en el tiempo, los rankings son básicamente inútiles más allá de la torpeza de fijarse en los resultados como si fuese una carrera de caballos y no preguntarse por el significado de éstos); así el (no) diálogo nacional sobre los resultados peruanos en PISA ha estado marcado por ser los últimos, o penúltimos, cuando esa posición relativa no da ninguna información que tenga algún significado.

Creo que estos son los puntos básicos que todos deberíamos conocer sobre esta nueva entrega de PISA. Sinceramente, debo insistir en que ya es tiempo de salirnos de esta prueba que no nos dice gran cosa y buscar alternativas (que existen) que son más ricas en cuanto a la información que nos pueden dar y que, sobretodo, no tienen la carga política (en el mal sentido de la palabra) ni los efectos perversos sobre la política y prácticas educativas que tiene PISA.