



VOL. 17, Nº 2 (mayo-agosto. 2013)

ISSN 1138-414X (edición papel)

ISSN 1989-639X (edición electrónica)

Fecha de recepción 24/04/2013

Fecha de aceptación 04/06/2013

DESDE EL ATRACTIVO PODER DE LOS DATOS DE PISA A LAS DESILUSIONES DEL BENCHMARKING. ¿DESAFÍO A LA EVALUACIÓN DE LOS SISTEMAS EDUCATIVOS?

From the Appealing Power of PISA Data to the Delusions of Benchmarking. Does that Challenge Any Evaluation of Educational Systems?



Marie Duru-Bellat

CNRS-OSC (Paris)

E-mail: marie.durubellat@sciences-po.fr

Resumen:

El objetivo de este artículo es mostrar y examinar el papel que pueden jugar los datos que aporta PISA en un marco más amplio relacionado con la evaluación de la calidad de los sistemas educativos. Para ello, algunas de las críticas y la resistencia dirigida a los datos ofrecidos por PISA en realidad tienen que ver con su mal uso en el proceso de la evaluación comparativa, así como, de una forma más amplia, con la evaluación de los sistemas educativos.

Palabras clave: PISA, benchmarking, rendimiento, evaluación, política educativa

Abstract:

The aim of this article is to show and discuss the role they can play PISA data provided in a broader framework related to the evaluation of the quality of education systems. To do this, some of the criticism and led resistance to the data provided by PISA actually have to do with their misuse in the process of benchmarking, as well as, more broadly, to the evaluation of educational systems.

Key Words: PISA, benchmarking, evaluation, achieving, educational policy.

1. Introducción

El objetivo de este artículo no es agregar ninguna crítica adicional respecto a los datos de PISA, a las observaciones técnicas ya existentes por parte de periodistas, responsables políticos y, a menudo de manera muy relevante, por los investigadores (por ejemplo, véase Goldstein, 2004, y los participantes de este documento). Más bien, vamos a discutir la importante cuestión sobre qué papel pueden jugar los datos PISA en el objetivo más global de la evaluación de la calidad de los sistemas educativos. Al hacer esto, vamos a sugerir que algunas de las críticas y la resistencia dirigida a los datos PISA en realidad tienen que ver con su mal uso en el proceso de *benchmarking*, así como, de una forma más amplia, con la evaluación de los sistemas educativos, lo cual es más discutible.

Esta perspectiva tiene importancia porque el desarrollo y el éxito de PISA se encuentran en un contexto global, con un fuerte énfasis en la educación, la financiación pública y la correlativa preocupación por la rendición de cuentas. En los países más ricos del mundo, el doble papel - económico y social - de la educación y la formación se ha venido reiterando. Esto es especialmente cierto para Europa, donde en 2000 se puso en marcha la Estrategia de Lisboa, cuyo objetivo era hacer de Europa “la economía basada en el conocimiento más competitiva y dinámica del mundo”. Tras ello siguieron los “objetivos de Lisboa”, la inclusión de un conjunto de objetivos con precisos *benchmarks* o puntos de referencia cuantitativos (Comisión Europea, 2007), como la participación cada vez mayor en la educación preescolar, la reducción del número de abandonos y del porcentaje de bajo rendimiento del estudiantado de 15 años de edad en lectura, el aumento del acceso a la educación superior, el desarrollo de la participación de los adultos en el aprendizaje permanente y la mejora de nivel de la población en todos los ámbitos educativos. Las cifras de los países de la UE se publican en informes periódicos; entre esta gran variedad de figuras, los datos de PISA han llegado a un público específico y ocupan un lugar importante en “*Education at a Glance*” y varias otras publicaciones de la OCDE.

2. ¿Por qué los datos de PISA son tan atractivos, pese a sus limitaciones?

El principal motivo del éxito internacional de PISA es sencillo: si la educación es considerada como un bien importante, en sí misma y como una herramienta estratégica de la competencia internacional, los datos de PISA proporcionan lo que se considera una evaluación objetiva de lo que “producen” los sistemas educativos. Para evaluar la cantidad de “capital humano” disponible en un país, se puede observar el porcentaje de individuos que han alcanzado un determinado nivel de educación. Por ejemplo, la OCDE (OCDE, 2009) publica periódicamente cifras sobre el porcentaje de adultos que han logrado una educación superior a la secundaria o terciaria, así como las tasas de escolarización y graduación por edades. Sin embargo, y especialmente cuando uno está interesado en las comparaciones entre países, limitarse a los números en relación al logro de una variedad de grados obtenidos por una generación dada no es suficiente, ya que no hay una correspondencia entre los grados de validez universal y los conocimientos y habilidades que certifican. Así, evaluar con precisión lo que los estudiantes han aprendido sobre una base comparable, es una herramienta mucho más valiosa. Por otra parte, esto también supone una preocupación legítima por parte de los gobiernos y los responsables políticos, ya que una parte importante de la financiación pública se destina a la enseñanza. ¿No sería sorprendente si los políticos no manifestaran ningún interés por la producción educativa?

De este modo, los datos de PISA resultan realmente atractivos, ya que dan una imagen concreta del rendimiento del alumnado de 15 años de edad en temas o ejercicios que son importantes para la vida diaria. El hecho de que lo que se evalúa sean supuestamente “habilidades para la vida” y no el puro conocimiento académico es especialmente atractivo, ya que la capacidad de hacer frente a la vida es el objetivo final de la educación, más que el dominio de los contenidos académicos. No importa qué críticas legítimas puedan plantearse con respecto a los datos de PISA, nadie podría argumentar que no resultan útiles para recoger información sobre lo que los jóvenes son capaces de realizar, a qué nivel de maestría y con qué disparidad. Por ejemplo, la mera información sobre el porcentaje de jóvenes de 15 años de edad con dificultades importantes en la comprensión de textos, es muy valiosa en sí misma.

De forma similar, independientemente de las limitaciones de los datos de PISA, sólo la demostración de que el promedio de rendimiento y las disparidades varían entre países con niveles similares de desarrollo y de financiación pública, es muy interesante. Además, es importante observar que la dispersión de los resultados varían en gran medida entre los países europeos: en algunos países, como España o Finlandia, el rendimiento estudiantil es mucho más homogéneo que en otros, como Bélgica o Alemania. Si bien esto puede parecer trivial, significa que el fracaso (o no) escolar, medido a los 15 años, no es una fatalidad, sino más bien se ve relacionado con factores claramente sociales y, por lo tanto, políticamente relevantes. Esto también es cierto en cuanto a las desigualdades sociales: mientras que las desigualdades de ingresos son similares en algunos países de Europa del Norte y del Este, las desigualdades sociales en el desempeño de los estudiantes evaluados por PISA resultan mucho mayores en los últimos (Duru-Bellat y Bydanova de 2009). Esto sugiere que algunos sistemas logran compensar mejor que otros la desigualdad existente en la sociedad en la que están inmersos, o al menos que algunas condiciones sociales, incluyendo ciertas características de los sistemas educativos, son importantes.

Los datos de PISA han sacado a la luz otros estímulos, incluso, una vez más, estando lejos de ser perfectos. Un ejemplo de ello es la prueba empírica que ofrecen sobre algunas ideas aplicadas actualmente en el mundo de la educación, como la posible (o incluso probable) relación entre la excelencia de la educación y su distribución, o lo que los economistas llaman un clásico intercambio entre la “eficiencia y la equidad”: dado que la búsqueda de la equidad requiere centrarse en los estudiantes más débiles, los mejores estudiantes sufren y el nivel de logro global es menor. En realidad, los datos de PISA demuestran que esto no es siempre así. Muestran que los países con una mayor proporción de estudiantes muy brillantes, con frecuencia son también aquellos en los que la brecha entre ellos y los más débiles, es más estrecha. Tal es el caso, por ejemplo, de los países escandinavos, mientras que muy pocos países tienen al mismo tiempo un porcentaje por encima de la media, y una gran variación de estudiantes medios (por ejemplo, Austria). De manera similar, la investigación sobre la base de los datos de PISA muestra que aquellos países con mayores desigualdades sociales también tienden a tener una mayor proporción de estudiantes débiles (por ejemplo, véase Duru-Bellat y Suchaut, 2005; Haahr, 2005). Por el contrario, los países con un alto grado de igualdad social tienen, en promedio, una mayor proporción de alumnos de alto rendimiento. Así que, con todo, no debe haber ninguna competición entre las políticas destinadas a mejorar la eficiencia media y las dirigidas a reducir la desigualdad social.

De hecho, en relación con la equidad, algunas características estructurales del sistema, como su grado de apertura y la diferenciación, pueden ser más importante. Por ejemplo, las comparaciones entre países sobre la base de los datos de PISA muestran que el

seguimiento temprano se asocia generalmente a una mayor desigualdad y un menor promedio de logros (Hanushek y Wößmann, 2006). Cualquier diferenciación curricular tiende al incremento de la desigualdad social porque los estudiantes privilegiados tienen más probabilidades de recibir el tipo de educación que contribuye a un mayor rendimiento (Perry, 2008). En términos generales, la investigación también muestra que en aquellos países con baja disparidad estudiantil, existe poca variación en el desempeño entre las escuelas, lo que significa que un cierto grado de calidad (definido aquí como el rendimiento alcanzado) está garantizado para todos los estudiantes, independientemente de sus centros de enseñanza. La magnitud de la dispersión académica también está relacionada con la cantidad de desigualdad social. En los países donde las escuelas no logran garantizar un rendimiento semejante para todos los estudiantes, las desiguales condiciones sociales tienen un mayor impacto en su rendimiento, lo que repercute en una mayor desigualdad social en los resultados educativos. En otras palabras, la homogeneidad en la calidad de la educación garantiza la igualdad educativa. Los datos de PISA también sugieren (con cierta cautela, ya que la composición escolar no está evaluada con exactitud en las muestras nacionales utilizadas) la importancia de los ingresos de los estudiantes, y producen un mayor rendimiento cuando más favorable resulta la composición socio-económica. Como dice Perry (2008, p. 83), "... en resumen, los resultados de la encuesta PISA sugieren que la desigualdad educativa puede resolverse mejor mediante una mayor similitud de las escuelas entre sí en términos estudiantiles, de currículum, y de recursos".

Sin ánimo de presentar una lista exhaustiva de todas las piezas de la información que se encuentran en los datos de PISA, se puede sostener que pese a ser imperfectos y cuestionables, siguen siendo útiles para poner de relieve las diferencias en los resultados educativos de los países. De tal forma, permiten debates que de otro modo se limitarían a un solo país, y dentro de un marco institucional, que es a menudo la manera normal y aparentemente obvia de pensar acerca de los problemas educativos, a riesgo de algún fatalismo. ¿La teoría de la reproducción social habría sido tan determinista si sus autores, Bourdieu y Passeron hubieran oído hablar de países sin importantes desigualdades sociales en las escuelas?

3. ¿Abusos y limitaciones de PISA, o del benchmarking?

Aquí nos centraremos en lo que parecen ser las desventajas principales de los datos de PISA, limitándonos a la forma de cómo se utilizan en las comparaciones entre países, en lugar referirnos a las limitaciones técnicas derivadas de la elección de temas específicos, imperfecciones en las muestras de estudiantes, problemas relacionados con las divergencias de conversión o las posibles diferencias culturales por las que los estudiantes entienden las preguntas de una forma u otra.

3.1. Algunos peligros de la obsesión por las clasificaciones

Desde un principio, la evaluación estudiantil se ha visto vinculada a las comparaciones internacionales. Esto es comprensible, ya que la evaluación no puede existir sin las comparaciones. Los primeros estudios internacionales (por ejemplo, el Primer Estudio Internacional de Matemáticas se puso en marcha en 1964 y, en términos generales, por la primera Asociación Internacional para la Evaluación del Logro Educativo, la IEA, además de otros estudios) tenían la clara intención de comparar países. A partir de la década de 1980, la

preocupación por la rendición de cuentas y los estándares se hicieron más y más convincentes, junto con la elaboración de unos indicadores estadísticos en todos los ámbitos de la vida económica y social. En 2000, la Comisión Europea puso en marcha sus dieciséis indicadores de calidad, destinados a ser los hitos del Método Abierto de Coordinación (OPC). Se les permitió la realización de un *benchmarking* regularmente, dando como resultado la identificación de aquellos países que funcionaban bien en determinadas áreas para que su experiencia y sus buenas prácticas pudieran, debido a una cierta presión de grupo, compartirse con otros. Dado que las comparaciones son la motivación última de todos estos estudios e indicadores, resulta necesario controlar la forma en que se hacen estas comparaciones.

El primer punto a destacar es que tanto en la educación comparada como en otras ciencias sociales, los resultados están supeditados a la muestra utilizada. Esto resulta evidente en cuanto a las clasificaciones, pero el grado de variedad en la muestra también es importante. Esto significa que no se puede estimar la influencia de una característica universalmente extendida mediante una muestra de países. Por otra parte, si en un análisis se combinan factores que varían poco en todos los países con otros que varían fuertemente, el peso de la segunda muestra será más evidente que el de la primera. Si se incluyen los países ricos y los muy pobres en la muestra de países estudiados, la riqueza del país puede aparecer como influyente en lo que respecta a las puntuaciones altas, lo cual no sería el caso con una muestra de países más homogénea. Asimismo, y como era de esperar, los factores contextuales significativamente asociados con el desempeño del estudiante pueden variar tanto para los países ricos y pobres, ya que muchos factores, como el nivel de formación del profesorado y el número de libros por estudiante, sólo entran en juego por debajo de un umbral mínimo.

Un segundo punto que limita en gran medida la validez de la comparación es el hecho de que aproximadamente (sólo) el 10% de la varianza total del resultado del estudiante en PISA se explica por la pertenencia a un determinado país (varianza entre países) (Haahr, 2005). Sólo este 10% puede estar relacionado con las diferencias entre los países y es a su vez explicable por las características contextuales nacionales y los factores organizativos de la escolarización. El restante 90% de la varianza se refiere a los propios países (varianza nacional). Esto significa que aquellos factores estructurales que varían de un país a otro, son de poca importancia en términos de las habilidades adquiridas por los estudiantes de 15 años, en comparación con las desigualdades de la familia, que causan un mayor impacto en las desigualdades entre los estudiantes. También puede ser que los países que participan en PISA compartan todas las características que afectan a la eficiencia (véase el punto anterior). Por lo tanto, la cifra del 10% no puede considerarse válida a escala mundial, y es probable que si se tomaran en cuenta aquellos países que varían en un grado aún mayor, en términos de los sistemas educativos, de riqueza o de nivel general de educación, los factores estructurales aumentarían en importancia. De hecho, algunos estudios muestran que los factores de organización de la escolarización pesan más en el rendimiento de los estudiantes en los países pobres que en los ricos (Heyneman & Loxley, 1983). Esto nos lleva a subrayar una vez más dos puntos fundamentales de la metodología de las comparaciones internacionales: 1) lo único que siempre evaluamos es aquello que varía, y 2) las relaciones identificadas dependen de la muestra del país que se está utilizando (es decir, variedad de la muestra).

Otro punto importante en relación a los datos de PISA, a menudo descuidado por los medios de comunicación (y también por los responsables políticos) se refiere a la inevitable falta de precisión de las cifras calculadas utilizando una muestra. Si se tiene en cuenta la matriz estadística de la varianza ligada a la muestra observada, las diferencias entre los

países similarmente desarrollados son muy limitadas. Sin embargo, cada clasificación tiende a magnificar las pequeñas desigualdades.

3.2. ¿Qué se evalúa?

Otro problema y quizás el más importante cuando se comparan países, se refiere a la naturaleza de lo que se evalúa en las puntuaciones de los mismos. Al igual que cualquier prueba, los datos de PISA se centran necesariamente en ciertas habilidades. Esto puede ser considerado como un activo, ya que requiere la puesta en marcha de los objetivos del sistema educativo, que a menudo se expresan en forma muy vaga. Sin embargo, aunque estos objetivos globales pueden ser consensuados, es mucho menos frecuente en el caso de las llamadas “habilidades para la vida”, especialmente cuando están definidas a nivel internacional, ya que tales habilidades pueden diferir de un país a otro.

En términos más generales, las comparaciones se basan en una selección de indicadores, lo que equivale a dar prioridad a determinados objetivos (por lo general, más fáciles de evaluar); dar prioridad a las puntuaciones medias académicas en el *ranking*, puede llevar a dejar de lado las actitudes de los estudiantes. Sin embargo, ningún país considera el dominio de ciertas habilidades como el objetivo único y último de la educación. En realidad, los datos de PISA también incluyen algunas evaluaciones de la opinión de los estudiantes sobre la vida escolar y algunas actitudes que se consideran productos colaterales a la escolarización (como el placer de trabajar en equipo). En ese sentido, los países han demostrado ser muy diferentes; en realidad, no hay una correlación entre los resultados académicos y el grado de satisfacción de los estudiantes con respecto a la escuela o a las actitudes “positivas” hacia los demás (Duru-Bellat et al, 2008.). En consecuencia, existe un cierto margen de elección; por ejemplo, algunos países “prefieren” conseguir que los estudiantes estén satisfechos con su vida escolar y aprendan a cooperar con sus compañeros, aunque tengan un rendimiento académico menos brillante.

Del mismo modo, la obsesión de clasificar a los estudiantes en base a la media de las puntuaciones puede llevar a descuidar la dispersión de las mismas. Sin embargo, se puede plantear si la clasificación debe basarse en el nivel académico de los estudiantes promedio, el 10%, o en los más débiles. En algunos países, se considera preferible tener un nivel de rendimiento medio moderado y homogéneo, mientras que otros países prefieren una situación inversa, con un nivel medio alto y mayores disparidades.

Por otra parte, para una variedad de razones, no pueden limitarse las mediciones a la edad de 15 años. En primer lugar, en algunos países, algunas la selección se produce antes, con el resultado de que tanto el nivel medio y la desigualdad entre los “supervivientes” son engañosas y no son directamente comparables con otros países. Además, en todos los países, los estudiantes pueden tener una desigualdad de oportunidades a lo largo de sus estudios, lo que significa que su nivel académico a los 15 años puede ser más o menos importante para su futuro. Desde una perspectiva más amplia, nos podemos plantear, con respecto a los economistas, a qué llaman eficiencia interna versus externa; es decir, a la hora de evaluar y comparar los sistemas educativos, ¿debemos dar prioridad a los estudiantes que lo hacen mejor o a aquellos que están mejor preparados para acceder al mercado laboral y experimentar la vida adulta? Volveremos a este tema más adelante.

Estas decisiones importantes, como la mayoría, deben originar un amplio debate, ya que ninguna de ellas se toma, a priori, mediante consenso. El problema central del

benchmarking es que los *benchmark* o puntos de referencia se establecen con los datos más fácilmente disponibles. Los datos de PISA son igualmente criticables, ya que utilizan una definición específica de lo que se considera valioso en un sistema educativo. Y esto generalmente se hace como si fuera obvio, mientras que en realidad no existen indicadores, incluso algunos indicadores económicos tradicionales como el PIB, parecen incuestionables ... Se puede sostener que una lista de indicadores y sus *benchmark* o puntos de referencia no equivalen a una evaluación del sistema del cual puedan extraerse consideraciones políticas. Sin embargo, no hay que confundir la hierba (es decir, la evaluación) con la maleza (es decir, los *benchmark*).

4. Evaluar: un proceso mucho más complejo que el benchmarking

El *benchmarking* es un medio para un fin; pretende ofrecer alguna orientación para las políticas educativas. En las ciencias sociales, las comparaciones siempre han sido consideradas como un sucedáneo de la experimentación, y la educación comparada se basa en esta idea. ¿Quién sostendrá que no es legítimo tomar algunas ideas prestadas de nuestros vecinos, aunque a menudo sea arriesgado y difícil?

4.1. ¿Tomamos prestados algunos elementos para aumentar la eficiencia?

Incluso si es preferible tener cifras antes que impresiones o ideas (sobre todo en lo que respecta a los temas educativos, porque todo el mundo tiene ideas!), la interpretación comparada de las cifras nunca es sencilla. La interpretación requiere hacer hipótesis sobre las causas de las variaciones observadas entre los países. Si vemos que los estudiantes de algunos países tienen un alto rendimiento y algunas características educativas, se tiende a dar un salto desde la correlación a la causalidad, y a “explicar” la primera con la segunda. Sin embargo, esto es una empresa arriesgada cuando se utilizan datos de PISA, por varias razones.

En primer lugar, los datos de PISA son transversales. Para los estudiantes de 15 años, esas diferencias son el resultado acumulativo de las articulaciones y los planes de estudio, el uso de las estructuras educativas y los tipos de relaciones entre profesores y alumnos que han prevalecido a lo largo de su escolaridad, además de todas las experiencias de aprendizaje desde el nacimiento hasta la actualidad. En otras palabras, los datos no sólo reflejan lo que el alumnado ha adquirido durante la escolaridad, sino también lo que han aprendido de su familia, sus compañeros, los medios de comunicación, la vida cotidiana, etc., y las capacidades evaluadas pueden ser más o menos dependientes de lo aprendido en la escuela. En realidad, las encuestas transversales de la IEA, como PISA, no permiten la evaluación de una eficacia real de los sistemas educativos, lo que requeriría datos longitudinales.

Otro punto es que la eficiencia educativa, probablemente se basa en una combinación de factores en lugar de uno sólo. Con demasiada frecuencia, los países de la OCDE al igual que otros tratan de sacar conclusiones de las correlaciones observadas entre una de las características precisas del sistema educativo, y el rendimiento de los estudiantes. En ese caso, el riesgo se encuentra en un exceso de interpretación de la correlación, que puede aparecer si los países específicos se incluyen en una muestra y desaparecen una vez que se eliminan. Aún más importante, la correlación puede ser falsa ya que se puede producir por otras características estadísticamente vinculadas a la anterior. De manera más general, un

clásico escollo en la comparación internacional es atribuir demasiada importancia a un elemento aislado de su contexto. En general, es arriesgado confiar en aquellas correlaciones que respondan a una determinada actuación y característica del sistema educativo, ya que de lo que probablemente se trata, es de un conjunto de rasgos que constituyen la coherencia social de cada país. Idealmente, la comparación debe tener en cuenta todas estas estructuras. Ciertamente, esto va en contra del objetivo más o menos explícito de estas comparaciones, es decir, la identificación de “leyes” *supra-nacionales*; en cuyo caso, las relaciones entre el modo de organización de un sistema dado y un “producto” estudiantil, serían transferibles.

Esto también va en contra de la preocupación de los responsables políticos sobre la transferencia de aquellas experiencias exitosas implementadas en otros países. Esta transferencia a menudo resulta imperfecta porque lo que funciona en un lugar, con ciertos tipos de estudiantes y de profesores, no siempre tiene éxito en otro contexto. De nuevo, esto se debe a que probablemente se trate de un conjunto de rasgos. Idealmente, cualquier comparación debe implicar todas estas estructuras globales. Por otra parte, el experimento exitoso debe poder adaptarse a las particularidades locales y a su cultura; por ejemplo, una estrategia exitosa para tratar con la inmigración en un país, donde sea sólo una pequeña minoría y proveniente de países específicos, puede fallar por completo en un país donde su número u origen sea muy diferente. Esta es también la razón por la cual el “método abierto de coordinación” (OPC) de la Unión Europea debe considerarse con cierta cautela, al igual que aquellas políticas “basadas en la evidencia”.

Se trata de un viejo debate en la educación comparada. Se opone a lo que algunos investigadores (Fuller y Clarke, 1994) llaman un “mecanismo de política” - la identificación de determinadas acciones escolares que aumentan el rendimiento estudiantil sin importar el contexto - con los “culturalistas de aula”, que sostienen que dichas acciones están siempre condicionadas por el contexto social, las normas y los entornos de las aulas. Sin entrar en este debate, lo que hay que destacar es que las piezas de información aisladas no son suficientes para evaluar los sistemas mecánicamente. Los indicadores resultan una valiosa información, pero la evaluación es un tema diferente, que requiere de la combinación de los mismos y, sobre todo, una interpretación más cualitativa de su significado (Bottani, 2008).

4.2. La evaluación requiere un constante intercambio y elecciones significativas

Mientras que los indicadores proporcionan valiosas y necesariamente piezas parciales de información, la evaluación de la calidad de un sistema educativo (con el fin de extraer algunas lecciones políticas) requieren la aplicación de una perspectiva mucho más amplia, y no carente de valores. Así, limitarse a los datos de PISA es claramente imposible.

A cada paso, las consideraciones aparentemente técnicas se mezclan con las de carácter normativo. Consideremos, por ejemplo, la forma en que un país define la excelencia en la educación. En realidad, se deben especificar muchas cosas. El rendimiento del estudiante debe definirse: 1) cuando se produce (por ejemplo, la habilidad lectora en la enseñanza primaria), es decir, cuando las personas se encuentran todavía en la escuela, 2) más adelante en su carrera escolar, en base a los requisitos de la educación secundaria, o 3) cuando acaban la escuela y son los adultos, basándose en las “habilidades para la vida”. Otra pregunta: ¿la excelencia en qué? ¿En base a qué aspectos definitorios del estudiante se evalúa tal excelencia? ¿Debemos definir algunos temas fundamentales o hacer un promedio de una

variedad de ellos? ¿Hay que dar más peso (o ninguno) a algunos de ellos? Otro aspecto es que la educación, por una parte, tiene como objetivo impartir conocimientos formales y por otra, los valores y normas de comportamiento. Mientras que los primeros son relativamente fáciles de medir y comparar, en el segundo caso es más discutible. Ninguna de estas cuestiones son puramente técnicas.

Por otra parte, resulta imposible (o al menos discutible) definir la excelencia sin tener en cuenta la equidad. ¿La excelencia de quién? ¿Hay que tener en cuenta el nivel medio o el nivel alcanzado por el 10%? ¿Debemos centrarnos en las diferencias entre individuos o grupos, o en la proporción de la población que se encuentra por debajo de un umbral mínimo determinado, de acuerdo con una perspectiva rawlsiana? En lo que se refiere a la escuela, el contenido de este umbral mínimo puede ser identificado con una perspectiva a corto plazo y de carácter académico (por ejemplo, cada individuo tiene que alcanzar un nivel de lectura mínimo establecido para el final de la escuela primaria). Sin embargo, también se puede adoptar una perspectiva a medio plazo, mediante la definición de lo que es necesario en la edad adulta (para una “buena vida”, como Rawls sugiere. Para profundizar sobre aquellos investigadores que adoptan una filosofía rawlsiana, ver Benadusi, 2007). Todo esto nos podría llevar a argumentar a favor de los indicadores de umbral, como el nivel mínimo de rendimiento académico formal para todos los estudiantes (en PISA, por ejemplo, el porcentaje de la población de un país, está por debajo de un umbral definido). Esto sería un criterio de equidad en las sociedades democráticas, garantizando la igualdad entre la ciudadanía y los derechos humanos básicos.

Por último, pero no por ello menos importante, al evaluar la equidad a nivel de sistemas, es imposible no tener en cuenta tanto la dimensión de la producción y la forma en que se han producido estos resultados. ¿Hay que tener en consideración el volumen de los recursos públicos asignados a la educación o el volumen de los recursos que los estudiantes reciben realmente? ¿Estos recursos son sólo dinero, o también tiempo, contenido, asistencia pedagógica, etc.? Otro grupo de preguntas hace referencia a si la excelencia se define mejor teniendo en cuenta la distribución de los recursos entre los estudiantes (por ejemplo, el grado de concentración). Así definida, la excelencia sería una combinación del rendimiento promedio y la equidad: no se mantiene un sistema excelente si se reciben recursos muy abundantes, pero se asignan a una parte muy limitada de la población. Del mismo modo, no podría decirse que un sistema es excelente si genera grandes desigualdades en el rendimiento de los estudiantes, o si genera pequeñas desigualdades, pero a un nivel muy bajo de logro. De cualquier manera, obviamente estas opciones encarnan valiosas consideraciones.

Todo esto implica que, al evaluar la excelencia, es importante no limitarse a las mediciones de rendimiento estudiantil, hay que incluir las mediciones de las características del sistema, tales como la cobertura, la financiación (pública/privada) y las estructuras (seguimiento temprano/integral, tipo de estudiantes, grupos, etc.) Cuánto más pobre sea un país, y más limitados sus recursos, más importante es centrarse en las características del sistema; sin embargo, esto sigue siendo una regla general, ¡ya que los recursos son siempre escasos y la excelencia un objetivo que nunca se alcanza plenamente!

Sea cual sea el país, es una opción valiosa decidir en qué medida el sistema debería centrarse en relación a la equidad. La pregunta es ¿cuántos años de escolarización son el objetivo para alcanzar una cobertura universal? Aquí la consideración de la equidad prevalece. Sin embargo, en esta etapa, la eficiencia importa una vez más. Una principal preocupación es cuánta educación básica debe ser garantizada a todos los niños del país. Dado que los recursos son siempre limitados, esto requiere tener en cuenta la eficiencia.

Ofrecer muy poco a todo el mundo en aras de la equidad no tendría sentido si la cantidad que se da es tan pequeña que no resulta efectiva. La equidad no puede ser el único argumento, depende de si se logra a un nivel de rendimiento bajo o alto. Una vez más, los juicios de valor son inevitables.

Por otra parte, los límites y los riesgos vinculados al benchmarking son claros: la lógica de este *benchmarking* nos lleva a evaluar un sistema desde un punto de vista puramente instrumental, no sobre la base de sus principios o valores, sino únicamente sobre la base de sus resultados, medidos con los indicadores disponibles. Desde esa perspectiva, la única preocupación es la eficiencia, por lo que “todo vale”... Todos los sistemas se consideran comparables, y parece posible “pedir prestado” una característica de un país a otro y construir el sistema ideal, como si se estuviera jugando con piezas de construcción. La eficiencia es la justificación para cada opción, las cuales son de carácter técnico y definidas por expertos.

5. Conclusión

De todas formas, en nuestra opinión, los legítimos debates y críticas que se centran en el *benchmarking* y la inclusión de datos de PISA no deben llevarnos a renunciar a los procesos de evaluación de los sistemas educativos en función de sus resultados. Las decisiones y políticas educativas deben estar justificadas, y los resultados que producen -los conocimientos, habilidades, actitudes, etc.- son ingredientes cruciales. Se podría incluso llegar a criticar cualquier sistema que no diera mucha importancia a los cambios que la educación produce en los estudiantes, ya que educarlos también es hacer que cambien de cierta manera, lo que justifica la perspectiva anterior. Aunque existen algunos debates sobre este tema (por ejemplo, Olson frente a Slavin, 2004), ¿qué métodos alternativos existen para justificar las políticas educativas, en las que se toman tantas decisiones?

Sin embargo, se puede entender fácilmente la variedad existente de resistencias a tal evaluación externa. En cuanto a los profesores, pueden pensar, legítimamente, que una evaluación precisa de los resultados de sus acciones, proporciona un marco más sólido para su trabajo. Por el contrario, algunas voces sostienen que la evaluación externa es lo que hace que una mayor autonomía sea posible, permitiendo al profesorado experimentar con las prácticas de su elección, siempre y cuando se cumplan las metas de producción. Otros piensan que con el desarrollo del *benchmarking* y la evaluación basada en los resultados, la forma en que el sistema se regula está cambiando, otorgando más poder a expertos y científicos, mientras que los responsables políticos y otros agentes no tienen más opción que ejecutar lo que demuestra su investigación. Sin embargo, la movilización de los socios educativos con objetivos globales (y definitivos) se puede considerar más importante que la polarización con aquellos otros de menor alcance.

Un último punto se ha de tomar en consideración, relevante tanto para el profesorado como para los investigadores de la educación comparada. Cualquier medida o evaluación de resultados se basa en lo que existe en la actualidad, no importa cuán sofisticado sean los métodos de análisis. Por ejemplo, una práctica pedagógica puede parecer la mejor entre todas las que existen. Dicho esto, ¡tal vez se podría diseñar alguna mejor! Además, no es suficiente con indicar que una acción tiene un impacto. Es necesario estar razonablemente seguro de que esta acción es mejor que todas las demás, con una misma cantidad de recursos. La verificación requiere la realización de una variedad de experimentos y, por

supuesto, no renunciar a la imaginación pedagógica. Hacer tal cosa sería desastroso para la educación comparada. Sin embargo, aunque limitarse a la utilización de datos de PISA sea un completo fracaso y constituya una perspectiva no sociológica, ¿qué significaría dejar de lado los resultados del alumnado para la educación comparada?

Referencias bibliográficas

- Benadusi, L. (2007). Education equality indicators in the nations of the European Union. In R. Teese, S. Lamb & M. Duru-Bellat (Eds.), *Internatinal studies in educational inequality* (pp. 155-190). Dordrecht: Springer, vol.1.
- Bottani, N (2008). Le niveau d'huile, le moteur et la voiture : les enjeux d'une évaluation de la qualité de l'enseignement par les indicateurs. *Education et formations*, 78, 9-23.
- Duru-Bellat, M. (2004). *Social inequality at school and educational policies*. Paris: UNESCO.
- Duru-Bellat, M., & Suchaut, B (2005). Organization and context, efficiency and equity of Educational system. *European Educational Research Journal* 4(3), 181-194.
- Duru-Bellat, M., Mons, N. & Bydanova, E., (2008), Cohésion scolaire et politiques éducatives. *Revue Française de Pédagogie*, 164, 37-54.
- Duru-Bellat, M. & Bydanova, E. (2009), L'éducation nationale : l'illusion égalitaire. *Revue du CIEP*. Disponible en : www.ciep.fr/ries/colloque-2009.
- European Commission, (2007). *Progress towards the Lisbon objectives in Education and Training*, SEC, 1284.
- Fuller, B. & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom, tools, rules and pedagogy. *Review of Educational Research*, 64(1), 119-155.
- Fuchs, T. & Woessmann, L. (2004). What accounts for international differences in student performance? A Re-examination using PISA data, *CESifo Working paper* No. 1235. Disponible en: www.CESifo.de.
- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education*, 11, 319-330.
- Haahr, J.H. (2005). *Explaining student performance. Evidence from the international PISA, TIMSS and PIRLS surveys*. Danish Technological Institute Disponible en: www.danishtechnology.dk.
- Hanushek, E.A. & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 11, 63-76.
- Heyneman, S. & Loxley, W. (1983), The effects of primary school quality on academic achievement across twenty-nine high and low income countries. *American Journal of Sociology*, 88(6), 1162-1194.
- Marks, G.N., Cresswell, J. & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational research and Evaluation* 12(2), 105-128.
- OECD (2004). *What Makes School Systems Perform?* Paris: OECD.
- OECD (2007). *No More Failures. Ten steps to equity in Education*. Paris: OECD.

OECD (2009). *Highlights from Education at a Glance*. Paris: OECD.

Olson, D.R. (2004). The Triumph of Hope Over Experience in the Search for "What Works", *Educational Researcher*, 33(1), 24-26.

Perry, L.B. (2008). Using PISA to examine educational inequality, *Orbis Scholae*, 2(2), 77-86.

Slavin, R.R. (2004). Education Research can and Must Address "What Works" Questions, *Educational Researcher*, 33(1), 27-28.