

La más bella del reino: el mundo de la educación en alerta con la llegada de un príncipe encantador¹

Norberto Bottani

Consultor y analista de políticas de educación

Resumen:

Este artículo intenta ofrecer un panorama crítico de la historia de las evaluaciones internacionales a gran escala que comienza en los años cincuenta del siglo pasado. La IEA (Asociación Internacional para la Evaluación del Rendimiento Educativo) fue el organismo pionero, y la primera área investigada fue la de Matemáticas. En 1958 aparece el primer borrador de proyecto de evaluación comparada de rendimientos de los alumnos en diferentes sistemas de enseñanza. El objetivo principal de este tipo de investigación es el llegar a identificar los factores que pueden explicar los rendimientos de los sistemas educativos para poder contestar a la pregunta crucial de si existe un sistema de enseñanza mejor que otro y cómo podemos definirlo para poder aprender de los mejores.

La experiencia acumulada de investigación comenzada por la IEA demostró, por vez primera, que se pueden medir empíricamente los resultados de la educación y aprovechar sus conclusiones. El mayor logro de la IEA ha sido poner en marcha instrumentos que hicieran factible la comparación de datos sobre el rendimiento de los alumnos en sistemas de enseñanza diferentes. En torno a esta asociación se crea una comunidad científica internacional especializada en este campo.

Su monopolio sobre la evaluación del rendimiento escolar acaba con la intervención y desarrollo, desde Estados Unidos especialmente, de otras asociaciones de evaluación que, a medio plazo, hacen reducir la autonomía científica de la IEA. A partir de la entrada en escena de la OCDE en 1993, las políticas internacionales de evaluación toman un nuevo camino. Las características del proyecto PISA, que posee una coherencia diacrónica, se concentra en tres dominios y evalúa niveles de competencia y no de conocimiento curricular concreto, hacen de sus conclusiones una poderosa arma de influencia sobre las políticas educativas de muchos países.

Palabras clave: IEA, evaluación del aprendizaje, rendimiento de los alumnos, estudios comparativos, sistemas educativos, OCDE, PISA, políticas educativas.

Abstract: *The Most Beautiful Girl in the Kingdom: the Excitement in the Education Field with the Arrival of a very Charming Prince*

Revista de Educación, extraordinario 2006, pp. 75-90.

⁽¹⁾ Adaptación de un texto elaborado para el *Haut Conseil français de l'éducation de l'école*. En Bottani y P. Vergnaud: *Francia y las evaluaciones internacionales*. HCEE, 16 (2005), p. 168.

This report sets out a critical view of the history of large-scale international assessments, which started in the fifties of the last century. The IEA (International Association for the Evaluation of Educational Achievement) was the pioneering body while the first subject to be assessed was Mathematics. In 1958, the first draft on the comparative evaluation project of students' achievement in different education systems was provided. The main objective of this type of research was that of identifying the elements which can explain the achievements of education systems in order to answer the crucial question whether there is an education system better than others and how we can define it so as to learn from the best ones.

The gathered experience on research developed by the IEA came to prove, for the first time, that educational outcomes can empirically be measured and that its conclusions can actually be useful. The major achievement of the IEA has been that of launching a set of tools in order to make feasible the comparison of data on students' achievements in different education systems. As a result, an international scientific group specialised in this field was created.

Its monopoly on achievement assessment ended up with the intervention and development –especially from the United States– of other assessment associations which may reduce the scientific autonomy of the IEA in the mid-term. From the intervention of the OECD in 1993, international assessment policies adopted a new perspective. PISA project's features, which show a diachronic coherence and focus on three major domains while assessing competence levels –and not those of specific curricular knowledge–, turn their conclusions into a very powerful and influential weapon on the educational policies of many countries.

Key words: IEA, learning assessment, students' achievement, comparative studies, education systems, OECD, PISA, educational policies.

LAS EVALUACIONES INTERNACIONALES A GRAN ESCALA: UNA HISTORIA DE 50 AÑOS

Hace aproximadamente 40 años, en 1965, la IEA (*International Association for the Evaluation of Educational Achievement*, conocida también en francés bajo el acrónimo AIE, cuya traducción al castellano sería *Asociación Internacional para la Evaluación de los Rendimientos Escolares*) estaba terminando su primer gran encuesta internacional sobre el conocimiento de alumnos en Matemáticas, que más tarde sería designada como la encuesta FIMS (First International Mathematic Study). Este primer ensayo había sido preparado en encuentros científicos organizados en el seno de la UNESCO a lo largo del precedente decenio; desde 1952 en el Instituto Internacional de Educación de Hamburgo donde se reunían, una vez al año, los directores de los más importantes institutos de investigación en educación de aquel entonces. Los asuntos tratados en dichos encuentros se referían a temas relativos a diversos aspectos de la evaluación de los programas de enseñanza, como por ejemplo, la elaboración de exámenes o la construcción de métodos e instrumentos para medir en qué proporción eran alcanzados los objetivos de la enseñanza y de los programas escolares. En 1958, los participantes a la reu-

nión fueron invitados a discutir sobre la posibilidad de organizar evaluaciones comparadas de rendimientos escolares obtenidos en diferentes sistemas de enseñanza. Benjamín S. Bloom, de la Universidad de Chicago, elaboró un primer borrador de proyecto de evaluación comparada de rendimientos escolares en diferentes sistemas de enseñanza.

Nadie en aquella época tenía experiencia en este tipo de investigación y nadie podía decir con antelación si los estudios de este tipo eran viables o no. El reto científico que este proyecto proponía subrayar no era mínimo: en efecto, se trataba de identificar los factores que podrían explicar los rendimientos de los sistemas de enseñanza para poder contestar a una pregunta crucial: ¿existe un sistema de enseñanza mejor que otro? ¿Quién sería entonces la chica más bella del reino? Esta pregunta un tanto iconoclasta no ha sido formulada por los dirigentes políticos sino por los científicos; es en el seno de la comunidad de investigación donde se comprometieron a estudiar una batería de pruebas con el objetivo de llegar a designar el mejor sistema de enseñanza, aquél que proponga soluciones susceptibles de obtener las prestaciones de conjunto más satisfactorias. Los que deciden han entrado en escena mucho más tarde. Podemos preguntarnos con este propósito si los científicos reunidos alrededor de este proyecto han actuado de manera totalmente independiente o si han anticipado un deseo no formulado de los que deciden. Estas preguntas merecerían ser estudiadas sobre la base de una documentación apropiada. La evaluación de los sistemas de enseñanza con ayuda de grandes encuestas de masas sobre competencias y conocimientos de los alumnos ha sido por lo tanto concebida a lo largo de los años cincuenta. Desde entonces, durante unos cincuenta años, hemos acumulado una amplia experiencia en estadística y metodología en este dominio. La evaluación de alumnos a gran escala a nivel internacional se ha vuelto una especialización en sí misma, un campo científico enteramente estructurado, con sus reglas de funcionamiento, de legitimación, de control y de reconocimiento

La realización de encuestas sobre los aprendizajes de alumnos escolarizados en los diferentes sistemas de enseñanza ha modificado de manera sustancial los estudios comparados tradicionales al abrir la puerta a pasos fundados sobre metodologías empíricas que miden los aprendizajes para juzgar la eficacia de los sistemas de enseñanza. Durante décadas, en los mejores casos, los estudios de comparación se basaban en la confrontación de datos simples, como por ejemplo el porcentaje de las tasas de participación en los diferentes niveles de instrucción, las tasas de escolarización más allá de la enseñanza obligatoria, la prolongación de la duración de la escolaridad obligatoria... Lo adquirido por los alumnos no se tomaba en cuenta ya que en general se estimaba que era muy difícil, si no imposible, medir los resultados de enseñanza con medidas empíricas. El proyecto científico lanzado por los pioneros de la IEA ha permitido desmentir en parte estas creencias y mostrar que se podía hacer algo para salir del estancamiento de las auto-evaluaciones o evaluaciones subjetivas de naturaleza holística o impresionista.

EL PRIMER ESTUDIO PILOTO A GRAN ESCALA DE COMPARACIÓN DEL «ADQUISICIONES» RENDIMIENTO DE LOS ALUMNOS

El primer estudio piloto para realizar un test sobre la posibilidad de realizar evaluaciones a gran escala del rendimiento de los alumnos en los diferentes sistemas de enseñanza se desarrolló entre 1959 y 1961. Doce países participaron con una muestra no representativa de 1.000 alumnos habiendo cumplido los 13 años a principios del año escolar, independientemente del curso cursado. Por lo tanto, 12.000 alumnos fueron objeto de este test, lo que en aquel entonces, era una población considerable, teniendo en cuenta los medios técnicos existentes para el tratamiento de los datos². De entrada, dos problemas se imponen: el de los criterios de definición de la población en cuestión (las edades o el grado escolar) y el de la elaboración de una muestra representativa. En este primer ejercicio, los dominios objeto de test fueron la comprensión de la lectura, las Matemáticas, las Ciencias y la Geografía. A su vez, un test no verbal puesto a punto en Inglaterra, fue igualmente realizado por todos los alumnos. Ninguna limitación de tiempo era impuesta para realizar el test: los alumnos podían tomar todo el tiempo que necesitaban para contestar a las preguntas. El test era acompañado de un mini-cuestionario dirigido a los estudiantes y a los directores de los centros. Los datos recogidos en 1961 fueron tratados y analizados por Bob Thorndike de la Columbia University de Nueva York, y los resultados fueron publicados en 1962.

El objetivo principal de este estudio piloto era probar la posibilidad de construir instrumentos utilizables de manera uniforme en diferentes sistemas de enseñanza (como por ejemplo podía ser el sueco o el japonés) con el fin de recoger datos que se pudieran comparar entre ellos y susceptibles de ser tratados de forma homogénea. Muchos elementos, que posteriormente se vuelven a encontrar de manera regular en todas las encuestas internacionales, se probaron y pusieron en marcha en este primer estudio piloto.

Los resultados del estudio piloto han sido tan satisfactorios que el grupo de los promotores decidió programar posteriormente un estudio real a gran escala. El estudio piloto había permitido resaltar y mostrar los puntos débiles y las lagunas del proyecto. Podríamos entonces esperar que corrigiendo los defectos de concepción del estudio, sobre todo en el tema de la construcción de muestras representativas de la población de alumnos³, pudiéramos lanzarnos a una operación ambiciosa, jamás intentada anteriormente, con probabilidades de éxito,

⁽²⁾ Las técnicas de evaluación comparadas han sido desarrolladas a lo largo de los años treinta en Estados Unidos, como por ejemplo en la encuesta «Eight-year Study» (Ocho años de estudio) que se desarrolló entre 1936 y 1941 bajo el pilotaje de Ralph Tyler (véase *International Encyclopedia of Educational Evaluation*, Pergamon Press, 1990, pp. 170-171).

⁽³⁾ Para poder obtener muestras probabilísticas válidas, el inglés Gilbert Peaker elaboró un manual de muestreo, que desde entonces se ha vuelto un texto de referencia para todos los estudios sucesivos de la IEA.

obteniendo el beneplácito de la comunidad científica y, sobre todo, el apoyo de los políticos que deciden y sin el cual no hubiera sido posible encontrar los recursos financieros para realizar semejantes proyectos.

EL NACIMIENTO DE LA IEA Y EL NACIMIENTO DE UN MONOPOLIO DE INVESTIGACIONES SOBRE LAS EVALUACIONES DE LOS SISTEMAS DE ENSEÑANZA

Al concluir este estudio piloto, había quedado claro que operaciones similares sólo podían ser llevadas a cabo por una organización especializada en llevar este tipo de proyectos de gran envergadura, reuniendo las experiencias necesarias para coordinar la participación de muchos institutos de investigación, planificar trabajos, asegurar la financiación y la gestión, coordinar el calendario de las pruebas, supervisar la comparación de las etapas, verificar la uniformidad de las herramientas, el tratamiento de datos y de cálculos, y para acabar, asegurar el análisis coherente de los resultados. Así, en 1961 fue constituida una asociación para la evaluación del rendimiento de los alumnos (la IEA, *International Association for the Evaluation of Educational Achievement*), cuya sede fue establecida cerca del Instituto Internacional de la Educación de Hamburgo. El primer presidente de la IEA fue William Douglas Wall, que era también el presidente de la NFER (*National Foundation for Educational Research*) inglesa. Tras solamente un año de actividad, Wall dimitió y en su lugar fue designado el sueco Torsten Husén, que quedó al mando de la IEA durante más de 15 años, hasta el final de 1978.

El mayor mérito científico de la IEA ha sido poner en marcha instrumentos que han hecho factible la mayor comparación de datos sobre el rendimiento de los alumnos que frecuentan sujetos a sistemas de enseñanza diferentes. Alrededor de la IEA, se constituyó en aquellos años una comunidad científica internacional especializada en la evaluación a gran escala de los conocimientos de los alumnos⁴. Muchas críticas han sido dirigidas a los estudios de la IEA, imputándoles sobre todo objetivos que éstos no tenían, como por ejemplo los de establecer explicaciones de causas de las diferencias entre sistemas de enseñanza. En verdad, existe una cierta ambigüedad en las tomas de posición de los pioneros de la IEA sobre este propósito, y no podemos excluir que un objetivo latente de este tipo haya inspirado a una parte de los pioneros, y que algunos de ellos cultivaran un diseño político en un sentido progresista que contase con recoger pruebas elocuentes a favor de reformas de enseñanza. Sin embargo, la apuesta principal, en el momento del inicio de esta historia, era otra, y consistía en desarrollar pruebas sobre el rendimiento de los alumnos, cuestionarios sobre el entorno

⁽⁴⁾ La IEA ha tenido un papel clave en la constitución y el desarrollo de la disciplina a nivel mundial, pero el desarrollo del dominio de la evaluación a gran escala del rendimiento de los alumnos no sólo se debe atribuir al éxito de la IEA. Por ejemplo, el papel principal lo han tenido los científicos que en Estados Unidos desde 1964 han puesto en marcha y realizado el NAEP (*National Assessment of Educational Progress*).

familiar y sus estrategias de aprendizaje, instrumentos para entender la proporción de currículums realmente enseñados así como el tiempo real durante el cual, los alumnos en clase se encontraban en condiciones de aprendizaje y enseñanza. Desde este punto de vista, la IEA ha ganado seguramente la partida: sus estudios han estimulado la investigación de sistemas de enseñanza y ha posibilitado a nivel mundial la realización de investigaciones que han dado a luz a nuevas representaciones de sistemas de enseñanza y modificado la comprensión de su organización y funcionamiento.

EL FIN DEL MONOPOLIO DE LA IEA

La IEA ha sido durante unos 30 años la única organización mundial especializada en el dominio de la realización de encuestas internacionales de masas sobre el rendimiento de los alumnos. Este monopolio se rompió en 1988 tras una decisión de la Administración federal americana que, después de haber financiado durante décadas a la IEA, decidió encargar al *Educational Testing Service of Princeton* (ETS) la realización de una encuesta internacional sobre el rendimiento de los alumnos haciendo competencia a la de la IEA. Esta encuesta, designada como IAEP (*International Assessment of Educational Progress*) fue realizada sobre dos asignaturas, las Matemáticas y las Ciencias. Con esta encuesta, repetida en 1991, el ETS proponía demostrar que era posible salirse de los carriles impuestos por la IEA, evitar los estancamientos que ralentizaban considerablemente las encuestas de la IEA y concebir un tipo de encuesta a nivel internacional menos pesada, más rápida que el elaborado por la IEA.

Las dos encuestas IAEP fueron muy criticadas por la comunidad científica, la cual estaba, en este momento, compuesta casi en su totalidad por especialistas formados en el seno de la IEA. La virulencia de la crítica llevó a la Administración federal americana a renunciar a su proyecto y parar la financiación del programa IAEP. La polémica de estas dos encuestas trataba en particular sobre el hecho de que los ítems de las pruebas, sólo eran, en gran parte, una trasposición a nivel internacional de los ítems de las pruebas de NAEP (*National Assessment of Educational Progress*)⁵, es decir, el test utilizado en Estados Unidos y puesto en marcha por el ETS para evaluar anualmente el rendimiento de una muestra de alumnos americanos. Por eso, el IEAP medía las competencias y el rendimiento de los alumnos de otros sistemas de enseñanza tomando como referencia crite-

⁽⁵⁾ El NAEP es el programa de evaluación puesto en marcha en Estados Unidos en 1964 para observar y medir las mejoras de los alumnos americanos de cuatro grupos de edad. La primera encuesta sobre una muestra nacional se realizó en 1969. El presupuesto inicial del NAEP fue aproximadamente de dos millones de dólares americanos. Después de 2001, con la reforma escolar propuesta por el Presidente Bush, el NAEP se ha vuelto una pieza central del programa de mejora de enseñanza. Su presupuesto ha alcanzado y sobrepasado los 100 millones de dólares. Sin ninguna duda, el NAEP ha sido un laboratorio formidable de renovación y experimentaciones de nuevos enfoques en materia de evaluación. Véase: Jones, V. Lyle y I. Olkin (2004) : *The Nation's Report Card. Evolution and Perspectives*. Phi Delta Kappa Educational Foundation, Bloomington.

rios americanos, lo que permitía evitar largas discusiones sobre la adecuación de las pruebas con respecto a los currículos nacionales.

La demostración de fuerza probada por la Administración federal con el IAEP no tuvo éxito. Los Estados Unidos fueron obligados a dar marcha atrás y renunciar a su proyecto, pero la operación consiguió, sin embargo, debilitar a la IEA y arrojarla en brazos de Estados Unidos. Se puede decir que el final del programa IAEP ha sido una «victoria pírrica» de la IEA. En efecto, sacando partido de la lección de dicho ataque, la IEA fue obligada a acelerar el ritmo de las encuestas sin tener medios. Podía encontrarlos en Estados Unidos pero para recibirlos había que someterse a las exigencias de la potencia dominante. Durante los años noventa, la IEA ha programado y tratado tres grandes encuestas: una sobre la comprensión de la lectura en 1990-91, una sobre las Matemáticas y Ciencias en 1994-95, que ha sido a partir de entonces, la mayor encuesta realizada a nivel internacional, con la participación de unos 40 sistemas de enseñanza diferentes, encuesta repetida en 1997-98, y por fin la encuesta sobre educación cívica en 1999⁶. Lo nunca visto en la historia de la IEA, si pensamos que en el curso del decenio 1960-70, la IEA ha realizado una sola encuesta; en el periodo 1980-92, dos. Sin embargo, en un sólo decenio, el que va de 1990 a 2000, la IEA, en crisis, lleva cuatro encuestas de gran envergadura. Es el efecto de una cura de un caballo propiciada por Estados Unidos. La Administración federal americana había perdido la batalla con el IAEP, pero también había conseguido retomar las riendas y perseguir sus intereses de una manera muy elegante.

LA ENTRADA EN ESCENA EN LA OCDE

El cese de las encuestas IAEP no ha sido, por lo tanto, una victoria para la IAE, que ha sido obligada como acabamos de ver a modificar sus prácticas y adaptarse a las exigencias de su arrendatario principal. La estrategia de los Estados Unidos ha conseguido debilitar a la IEA haciéndole perder su autonomía científica y sobre todo al englobarla *coram populi* en la esfera de influencia de los Estados Unidos. De hecho, la puerta estaba abierta para un tercer actor, como es el caso específico de la OCDE, que entra en acción en 1993, cuando la organización se da cuenta de que nunca hubiera obtenido la información que necesitaba sobre los aprendizajes de los alumnos para completar su conjunto de indicadores. La antigua IEA no lo conseguía, la nueva IEA estaba demasiado bajo control de Estados Unidos para obtener la confianza de todos los países. En 1997, los ministerios de Educación de los países de la OCDE decidieron lanzar un ciclo de encuestas autónomo, de un nuevo tipo, sobre los conocimientos de los alumnos. En la primavera del 2000, unos 30 países participaron en el primer estudio del Programa Internacional sobre las Evaluación de los Alumnos (PISA, en sus siglas en inglés).

⁽⁶⁾ A estas encuestas conviene añadir la encuesta sobre las nuevas tecnologías en la escuela (IEA-SITES), realizada en 1998-99.

La OCDE ha obtenido, por lo tanto, lo que la IEA nunca había conseguido realizar en cuarenta años, es decir, no solamente atraer la atención de los que deciden sobre estas encuestas, sino reorientar las políticas de la enseñanza en muchos países; PISA se había vuelto un referente para justificar todo tipo de decisiones o reformas, (OCDE, 2004).

LAS NOVEDADES DE PISA

Sería falso afirmar que no hay solución de continuidad entre PISA y los estudios anteriores de la IEA. PISA, aunque retoma muchos conocimientos metodológicos puestos en marcha por la IEA, representa una ruptura con respecto a las encuestas anteriores. Las principales novedades de PISA, que desmarcan sus estudios de las de IEA, son las siguientes.

LA PERIODICIDAD TRIENAL DE LAS PRUEBAS

La primera novedad en el proyecto consiste en programar un ciclo de encuestas según una periodicidad fijada con el fin de construir series históricas de datos comparables sobre las competencias de los alumnos. PISA ha adoptado una periodicidad trienal. Las razones de esta elección no son claras, pero se puede suponer que el objetivo de la OCDE era demostrar no sólo que era posible organizar encuestas a gran escala a intervalos cercanos como lo había intentado hacer antes el ETS, sino que también tenía la capacidad de mantener este ritmo sobre un largo período, como es un período decenal⁷. Una periodicidad de tres años pone bajo presión a los sistemas de enseñanza; no se desentienden las administraciones; los directores de los centros, los institutos de investigación encargados de realizar las encuestas: todo el mundo debe adaptarse, aunque la cantidad de datos producidos cada tres años no sea explotada. Qué más da, mientras haya resultados, que los políticos que deciden, desean. Sin el acuerdo, el apoyo o la neutralidad benévola de los políticos que deciden, que por otro lado pagan los costes del estudio, la OCDE no hubiera podido realizar su programa. Otro factor, no sin importancia para el éxito, es el apoyo de una parte de la comunidad científica, que lejos de ser reacia de cara a un ritmo desenfrenado de recogida y producción de datos, «mantiene el tipo» y saca un beneficio no sólo científico sino también financiero participando en esta operación. Si hay reservas o reticencias, éstas no son visibles o no aparecen en los informes⁸.

⁽⁷⁾ Conviene señalar que la IEA ha adoptado también en los años noventa, por lo tanto, después de la crisis desencadenada por el IAEP, una modalidad periódica en la programación de encuestas. Así, aunque guardando el acrónimo TIMSS que significa «Tercera encuesta internacional sobre las Matemáticas y las Ciencias», cuando la encuesta se realizó en 1994-95, se trataba de la tercera encuesta de la IEA sobre este tema. Con la réplica de la encuesta en los siguientes años, hubo que deletrear de otra manera el acrónimo: la «T» y así se pasó del término «Tercera» al término «Tendencias» (*trends* en inglés).

⁽⁸⁾ Es interesante señalar que, en ciertos países, en el momento de presentar resultados de PISA 2000 y PISA 2003, se han verificado qué reformas de enseñanza habían vivido los alumnos que habían sido sometidos a las pruebas. Por ejemplo, en el momento de la presentación de resultados de PISA 2003, el Presidente de la Conferencia de los Jefes de Departamentos de Educación de regio-

LA COHERENCIA DIACRÓNICA DE LAS PRUEBAS

El principio de la repetición de un estudio en el mismo campo disciplinario no es en sí una novedad. La IEA había puesto en marcha un programa de estudios sucesivos en el área de las Matemáticas y en el dominio de la Lectura, ya que el interés de realizar comparaciones diacrónicas de resultados obtenidos por un mismo grupo de edad en el interior de un mismo sistema de enseñanza (pero a diferentes fechas) es evidente para todo el mundo. Sin embargo, con PISA, es la primera vez que se lanza un programa fijando de primeras la periodicidad de las pruebas. Los beneficios de la participación en PISA podrán entonces sumarse sólo si un sistema de enseñanza se compromete a lo largo de dos ciclos de encuestas, es decir en un periodo de 12 años. Por ejemplo, según el plan PISA, sólo después de nueve años serán evaluadas de nuevo con la misma profundidad analítica las competencias en la comprensión de la Lectura, dominio evaluado en la primavera de 2000, y que no será asignatura principal hasta el test de 2009. Esta estrategia permite a la OCDE asegurarse el compromiso de los países (los clientes!) por un largo periodo de tiempo, obligándoles a participar en la encuesta y a pagar esta participación durante 12 años para sacar en contrapartida el pleno beneficio de la participación al programa. Por supuesto, nos podemos retirar del programa, pero en este caso perderíamos la ventaja ofrecida por la posibilidad de realizar comparaciones a largo plazo.

CONCENTRACIÓN SOBRE TRES DOMINIOS

La tercera novedad del programa PISA es la elección de tres dominios probados de forma simultánea aunque con un orden de importancia diferente. Los tres dominios tomados en cuenta por la OCDE son la comprensión de la lectura, la cultura matemática y la cultura científica. Cada vez, los tres dominios son objetos de prueba, pero el sitio reservado en el test a cada uno de estos dominios es diferente. En el programa PISA se habla de dominios mayores y de dominios menores. Para cada encuesta, una de estas materias será evaluada al detalle, su estudio representará más ó menos dos tercios del tiempo de la duración total de las pruebas. También, y como ya era el caso en las encuestas IEA, los alumnos además tendrán veinte minutos para contestar a un cuestionario contextual sobre su motivación, su entorno familiar y cultural, sus compañías, sus intereses. El cruce de las respuestas en el cuestionario con los resultados obtenidos en el test va a permitir examinar los efectos de las diferentes estrategias de aprendizaje puestas en marcha por los alumnos según su perfil social, cultural y económico.

nes helvéticas, M. Stöckling, declaró que el puesto relativamente bueno de los alumnos suizos era la prueba de la eficacia de las reformas de la enseñanza realizadas en los años noventa.

UNA POBLACIÓN ESTABLE Y BIEN IDENTIFICABLE: LA DE LOS ALUMNOS DE 15 AÑOS

Otra originalidad del programa PISA consiste en la composición de la población evaluada, los jóvenes de 15 años, independientemente del curso al que asistan. Esto significa que en la muestra representativa de la población de alumnos de 15 años habrá alumnos de cursos escolares diferentes. PISA no verifica lo que se ha aprendido en un curso escolar. Es el nivel cultural de los jóvenes de 15 años en el país objeto de la encuesta. En los países de la OCDE la mayoría de los jóvenes de 15 años están en la escuela y por ello es relativamente fácil encontrarlos cuando se compone una muestra representativa de la población escolar. Una encuesta sobre los jóvenes de 15 años es menos cara: sabemos dónde están los jóvenes y podemos hacerles el test en la escuela. Este paso presenta sin embargo un inconveniente: en los sistemas de enseñanza donde se practica a gran escala el repetir curso, encontraremos en la muestra de la encuesta una proporción más o menos importante de alumnos de 15 años con retraso de un año, o quizás dos con respecto a la clase teórica donde se encuentra la mayoría de los alumnos de 15 años que tiene una escolaridad regular. Estos alumnos pasan por lo tanto el test sin haber realizado toda la escolaridad obligatoria, lo que puede presentar una desventaja con respecto a una muestra de jóvenes de 15 años que hayan sido escolarizados en un sistema de enseñanza en el cual no se practica casi nunca el repetir curso.

La opción de dar preferencia al grupo de edad como criterio de composición de la muestra evaluada en vez del curso, presenta otra particularidad que es contradictoria con uno de los objetivos que pretende PISA. En efecto, la elección del grupo de edad de 15 años se justifica por el hecho de evaluar las competencias al final de la escolaridad obligatoria, estimando que la mayoría de los jóvenes de los países de la OCDE completan su escolaridad obligatoria a la edad de 15 años. Pero, esta suposición no es del todo correcta. En efecto, si quisiéramos verificar el nivel de competencias adquirido en el interior de un sistema de enseñanza determinado por el conjunto de alumnos acabando la escolaridad obligatoria, no se hubiera tenido que elegir el grupo de edad para constituir la población evaluada, sino el último curso de la enseñanza obligatoria. Esta opción hubiera implicado la constitución de una muestra de alumnos compuesta de individuos con edades distintas. Esta opción, más compleja de gestionar, no ha sido adoptada por la OCDE.

PRUEBAS CENTRADAS EN LAS COMPETENCIAS Y NO EN EL CURRÍCULUM ESCOLAR

PISA no propone evaluar lo que se aprende en la escuela con respecto a los programas que los docentes deben o son invitados a enseñar. Esta operación se había revelado muy ardua en el transcurso de las encuestas de la IEA. Con la encuesta TIMSS (*Third International Mathematics and Science Study*), la IEA había encargado una encuesta exhaustiva de programas de enseñanza de Matemáticas con vistas a interpretar los resultados de las pruebas sin conseguir

establecer la conjunción entre programas, pruebas y resultados⁹. La estructura y la organización de los programas de enseñanza es un rompecabezas para los «comparativistas». PISA resuelve las dificultades apartándolas. Las pruebas de la OCDE no se centran en los programas sino en las competencias o aptitudes juzgadas indispensables para llevar una existencia autónoma e independiente en las sociedades democráticas con una economía de mercado¹⁰.

Por estas razones, las empresas en competencia con la IEA en el campo de las evaluaciones internacionales de competencias han seguido otro camino. Las dos encuestas del ETS conocidas bajo el acrónimo de IAEP 1 y 2 (*International Assessment of Educational Progress*) realizadas en 1988 y 1991 han sido concebidas para evaluar los conocimientos en dos disciplinas, las Matemáticas y las Ciencias. Sin embargo, para compensar la unión entre pruebas y programas nacionales de enseñanza, el ETS ha adoptado y parcialmente adaptado los ítems del test de programa de evaluación federal americano NAEP (*National Assessment of Educational Progress*). Aún así, el problema de la concordancia entre test internacional y programas de enseñanza estaba resuelto teniendo como único referencial el programa de enseñanza en vigor en sólo uno de los sistemas de enseñanza. Tomar en cuenta las especificaciones de los programas de enseñanza en vigor en los otros sistemas de enseñanza sólo ha ocurrido en el momento de la interpretación de resultados, ponderando éstos en función de los programas.

Cuando la OCDE toma la decisión de realizar evaluaciones a gran escala a nivel internacional con un ritmo muy constante, también se encuentra el problema causado por las diferencias existentes entre los programas de enseñanza. Para proceder de manera rápida, había que evitar mezclarse con este tipo de problemas. El camino que siguió el ETS no era practicable, porque no era políticamente correcto; desde entonces, sólo quedaba una solución: la de liberarse totalmente, en la elaboración de las pruebas, de todos los programas de enseñanza.

⁽⁹⁾ El análisis de los programas de enseñanza ha sido objeto de una atención particular en el caso de la encuesta TIMSS que se desarrolló entre 1994-95. En esta ocasión, un grupo específico de trabajo dirigido por William H. Schmidt de la Universidad de Michigan analizó los programas de enseñanza oficiales y los manuales de Matemáticas en uso en los sistemas de enseñanza que participaron en este estudio. El proyecto SMSO (*Survey of Mathematics and Science Opportunities*) ha permitido constituir un banco de datos sobre programas de Matemáticas y los manuales escolares utilizados en clase para la enseñanza de las Matemáticas. Véase la obra *Characterising pedagogical flow: an investigation of mathematics and science teaching in six countries*, publicada por Dordrecht, Kluwer Academic Publishers en 1996. Desgraciadamente, este análisis de los programas versaba sólo sobre un pequeño número de sistemas de enseñanza y, sobre todo, se ha desarrollaba en paralelo a la encuesta TIMSS y no la ha precedía como hubiera sido lógico esperar para construir una evaluación orientada según el currículum. El problema de adecuación entre sistemas de enseñanza y evaluaciones internacionales ha constituido un obstáculo permanente para la IEA.

⁽¹⁰⁾ Por ejemplo, los resultados insuficientes obtenidos por los alumnos de un sistema de enseñanza en el cálculo algebraico o de geometría espacial pueden explicarse por el hecho de que estos capítulos de la enseñanza de las Matemáticas no habían sido tratados en la escuela o ni siquiera estaban incluidos dentro del programa de enseñanza.

Es el camino seguido por la OCDE en PISA, ; se ha renunciado a medir lo que los alumnos aprenden en la escuela (por lo menos en teoría), y se mide, por el contrario, el nivel de competencias que es considerado a los 15 años como indispensable para vivir en sociedades democráticas y con economía de mercado. De hecho, las pruebas PISA han sido concebidas como PRUEBAS «curriculum free», es decir pruebas neutras con respecto a programas de enseñanza. El test PISA tiene como objetivo evaluar la presencia o la ausencia de aptitudes juzgadas «esenciales» para una vida de adulto¹¹. Según los mismos términos de la OCDE, «se trata de la aportación más importante y ambiciosa del programa OCDE/PISA»¹².

El programa PISA se libera de límites impuestos con la necesidad de encontrar un denominador común constituido por el contenido específico de la enseñanza impartida en las escuelas de los países participantes. Este enfoque de la evaluación es justificado por la OCDE por la exigencia de valorar el capital humano definido como «los conocimientos, cualificaciones, competencias y otras cualidades que posee un individuo y que interesa al bienestar personal, social y económico». La ambición de la OCDE es llegar a medir de manera apropiada el stock de capital humano de un país. PISA marca así un giro con respecto a los estudios anteriores a gran escala. Consagra el hecho de que la evaluación de conocimientos y contenidos escolares no es en sí suficiente. El programa PISA transmite un nuevo mensaje, esto es, que la eficacia de los sistemas de enseñanza no se mide con respecto a la enseñanza y a la adquisición de conocimientos, sino por evaluaciones centradas en el desarrollo del capital humano. Los parámetros para evaluar un sistema ya no son internos al sistema, sino externos.

LA POLITIZACIÓN DEL PROGRAMA DE EVALUACIÓN

Con PISA se realiza el relevo entre la comunidad científica que había llevado de manera más o menos autónoma, de manera aislada, los estudios a gran escala durante unos 30 años y la comunidad de políticos que deciden, que manejan los asuntos y dictan o imponen sus prioridades. La transición es clara: PISA tiene recur-

(11) OCDE: *Medir los conocimientos y competencias de los alumnos. Un nuevo ámbito de evaluación*. París, 1999. En los documentos de la OCDE se menciona un nuevo referencial: las competencias necesarias para tener éxito en su vida (*for a successful life*), que es menos preciso que los conocimientos o competencias fijadas en los programas de enseñanza.

(12) Posteriormente, la OCDE ha patrocinado un proyecto financiado por Estados Unidos y la Confederación Helvética cuyo objetivo era definir componentes que permitan crear en el interior de cuentas nacionales una cesta de bienes educativos que hubieran tenido que permitir elaborar un índice PPA (poder y paridad de compra) para educación, índice cuya necesidad se ha hecho notar al producir indicadores financieros de «*Regards sur l'éducation*» que son comparables sólo al ponderar los valores nacionales en función del coste de vida existente en los diferentes países. Sin embargo, el índice PPA producido y utilizado por la OCDE se ha establecido sobre la base de cálculo del coste de la cesta de bienes esenciales en el cual no existen bienes relativos a educación. Es para identificar estos bienes educativos por lo que el proyecto DESECO (Definición y Selección de Competencias Claves) ha sido lanzado. Posteriormente, el proyecto ha sido la apuesta para la elaboración de evaluaciones de encuestas sobre la literatura de los adultos (*Project ALL*) y de alumnos de 15 años (proyecto PISA).

sos y ha podido imponerse porque está al servicio de lo que esperan los que deciden y tienen la responsabilidad de la conducción dirección y de la regulación de los sistemas de enseñanza y de los aparatos burocráticos que los gobiernan.

Cuando, en otoño de 1997, los gobiernos de la OCDE decidieron lanzar PISA y repartir entre ellos los costes de su financiación, una nueva época empezó en la historia de las evaluaciones internacionales a gran escala. En efecto, por primera vez, los gobiernos han asumido la responsabilidad de programar evaluaciones a gran escala con el objetivo de obtener informaciones estimadas necesarias para regular los sistemas de enseñanza y no sólo para objetivos de conocimientos propuestos por la comunidad científica. La decisión ha sido tomada sin prestar gran atención a las repercusiones a medio y largo plazo de este proceso. Sin embargo, desde hace mucho tiempo, sabemos que la realización de evaluación en las clases y sobre los sistemas de enseñanza provoca comportamientos de adaptación que sólo conocemos en parte. Las evaluaciones ponen bajo presión al sistema y a sus actores. Muchas pruebas, señales y declaraciones indican que PISA no ha podido tampoco evitar este riesgo. Se entiende que la repetición cíclica de evaluaciones previstas en el programa va a empeorar, a medida que pasen los años, las reacciones de adaptación y de preparación a la evaluación, distorsionándose así, en un grado que no podemos medir todavía, los resultados de las evaluaciones y las comparaciones.

Subyacente a esas preguntas se sitúa el conflicto clásico que opone exigencias y prioridades científicas de un lado e imperativos políticos del otro, entre conocimiento y cambio, entre objetivos teóricos y control del servicio. Al financiar PISA, las autoridades escolares responsables de los sistemas de enseñanza esperaban no sólo obtener de manera rápida datos sobre los efectos de políticas educativas y de reformas escolares, sino, sobre todo, controlar la eficacia de los sistemas, y obtener informaciones sobre diversas formas de regulación de la enseñanza. Esta apuesta ha sido ganada de manera arrolladora, si nos ceñimos a las repercusiones de la encuesta PISA sobre las políticas nacionales de educación de muchos países.¹⁵

LA FLEXIBILIDAD DEL PROGRAMA

Una última particularidad de PISA es su flexibilidad: los sistemas de enseñanza que participan en el programa pueden añadir al test principal dos partes opcionales (cuestionario o test). En el primer lanzamiento de PISA en el año 2000 fueron propuestos a los países, dos cuestionarios opcionales, uno sobre competencias transversales y otro sobre competencias informáticas. El primero de estos

⁽¹⁵⁾ Véase por ejemplo en Francia los informes redactados por el *Haut Conseil de l'évaluation de l'école*, como el informe de M. T. Ceard; M. Remond; M. Varier: «L'application des compétences des élèves et des jeunes en lecture et en écriture et l'évolution de ces compétences dans le temps», diciembre de 2003, o el informe de M. Salines; P. Vrignaud: «Apprécier et certifier les acquis des élèves en fin de collège: diplôme et évaluations-bilans», junio 2001.

dos cuestionarios fue preparado en el seno del proyecto INES de la OCDE-CERI por una red de científicos que se preguntaban sobre los productos de la educación escolar. Tras aproximadamente siete años de trabajo, este grupo ha conseguido incluir un cuestionario en el paquete de instrumentos propuestos por el programa PISA. Una acción similar ha llegado a la elaboración del cuestionario sobre las competencias informáticas: 26 países de 32 han utilizado los cuestionarios sobre las competencias transversales y 20 de 32 el del uso de los ordenadores y la disponibilidad de ordenadores en las escuelas y en casa, así como la apreciación que los estudiantes daban sobre sus competencias informáticas. Estas partes han sido negociadas a nivel internacional; otras han sido puestas en marcha en el seno de un sistema de enseñanza o en colaboración con ciertos sistemas. Por ejemplo, los sistemas de enseñanza germano hablantes han estudiado juntos un cuestionario sobre los docentes que no se ha llevado a cabo. Cuestionarios y módulos complementarios, como por ejemplo el cuestionario sobre estrategias de aprendizaje de los alumnos, han sido muy explotados en la interpretación de los resultados del test principal. Aunque sea difícil incluir en un programa de evaluación de por sí ya muy cargado, otras partes que alargan el tiempo de la duración de la aplicación con consecuencias no previsibles sobre la motivación y la concentración de los jóvenes y, por tanto, sobre las posibles distorsiones en los resultados, esta eventualidad abre caminos para recoger otras informaciones sobre los centros. Nada impide que el programa PISA se vaya modulando si no es por su rigidez y su duración, que hacen de él un programa omnívoro que no deja con facilidad espacios para explotaciones complementarias.

CONCLUSIÓN: LA FIABILIDAD DE ENCUESTAS INTERNACIONALES A GRAN ESCALA

Muchas críticas siguen formulándose con respecto a las encuestas internacionales y en particular respecto a PISA. La explotación realizada sobre el plan político y por los medios de los resultados de PISA no suscita confianza en los medios educativos. También hay una serie de problemas metodológicos que resolver y esclarecer como, por ejemplo el modelo de medición utilizado, las condiciones de validez de los ítems del test, la reducción de distorsiones culturales, las cuestiones de traducción, etc. Han sido realizados adelantos considerables en el transcurso de los últimos cuarenta años sobre el plan metodológico y de organización. La calidad de las últimas grandes encuestas internacionales como PISA, PIRLS, TIMSS (*Trends in Mathematics and Science*) es admirable. Estos estudios son una mina de información sobre los sistemas de enseñanza. Por otro lado, la exigencia de la comparación revela sus fallos. Sin embargo, son insustituibles, ya que no hay otros medios para documentarse sobre las prestaciones de servicios de enseñanza para realizar comparaciones que digan algo válido sobre el estado de salud de un sistema, produciendo datos empíricos que pueden ser discutidos, puestos a prueba o rechazados sin más.

El reverso de la moneda es la tentación de la clasificación para designar a «la chica más bella del reino». Sin embargo, este espejo que nos muestran los estu-

dios internacionales no es perfecto. Es un espejo que deforma. Sólo un esfuerzo considerable y crítico sobre el plan científico podrá ayudar a corregir sus defectos. Más que nunca, debemos tener un debate democráticamente abierto y una discusión transparente sobre la naturaleza y los límites de estas encuestas. Su calidad es indiscutible, la calidad de las evaluaciones es evidente, las técnicas de muestreo se han perfeccionado, el tratamiento estadístico ha avanzado, sobre todo en lo que se refiere a la metodología de comparaciones internacionales. Conocemos mejor los límites y las trabas de estos estudios y empezamos a dominarlas mejor. Por lo tanto, una conclusión se impone: si queremos controlar los efectos de estas estrategias y dominar estas operaciones, poder pronunciarnos e influir en ellas, hay que participar, ya que sólo así haremos progresar la ciencia de la educación pero también podremos adquirir un conocimiento íntimo de mecanismos, de luchas de poder, de grupos de presión que bajo bastidores forjan estos estudios. Este conocimiento es indispensable para actuar con eficacia, sin cometer graves errores que restarían cualquier credibilidad a las quejas y a las tomas de posición que habrá que formular ante la comunidad científica.

Traducción: Claudine Corrocher

REFERENCIAS BIBLIOGRÁFICAS

OECD (2004): *What Makes School Systems Perform? Seeing School Systems Through. The Prism of PISA*, París.

SALGANIK, L.; RYCHEN, D. (eds.) (DeSeCo) (2001): *Defining and selecting key competencies*. Hogrefe & Huber Publ, Seattle.

SIMONE RYCHEN, D.; HERSH SALGANIK, L. (2003): *Key Competencies for a Successful Life and a Well Functioning Society*. September.

HAUT CONSEIL DE L'ÉVALUATION DE L'ÉCOLE (2001 Y 2003):

— CEARD, M. T.; REMOND, M.; VARIER, M. (2003): «La aplicación de competencias de los alumnos y los jóvenes en la lectura y escritura y la evolución de estas competencias en el tiempo», (Informe, diciembre).

— SALINES, M.; VRIGNAUD, P. (2001) «Apreciar y certificar los conocimientos de los alumnos al finalizar el colegio: diplomas y evaluaciones-balances», (Informe, junio).