

# La evaluación como instrumento de política educativa

JORGE CALERO

*Universidad de Barcelona e Institut d'Economia de Barcelona*

ÁLVARO CHOI

*Universidad de Barcelona e Institut d'Economia de Barcelona*

*Recibido: Abril 2012*

*Aceptado: Junio 2012*

## Resumen

El objetivo del artículo consiste en describir y valorar los efectos de la evaluación de competencias, una de las piedras angulares de las políticas educativas actuales. Se describe en el artículo la diversidad de funciones alternativas atribuidas a los sistemas de evaluación y se discute el papel conferido a la publicación de rankings de centros. Una aproximación al «mapa» actual de las evaluaciones educativas en España permite comprobar el efecto hegemónico de las evaluaciones de la IEA y la OCDE. En el artículo se valoran las consecuencias (deseadas y no deseadas) de la aplicación de sistemas de evaluación de competencias.

**Palabras clave:** evaluación educativa, PISA, rendimientos de la educación, competencias.

**Clasificación JEL:** H.52, I.21.

## Abstract

In this paper we aim at describing and discussing the effects of competence evaluation, one of the cornerstones of current educational policies. We describe a diversity of alternative functions which are conferred on the evaluation systems. The role played by the publication of rankings is discussed. The hegemonic effect of IEA and OECD's evaluations can be confirmed by reviewing the current «map» of educational evaluations in Spain. In the paper the (intended and unintended) consequences or competence evaluation systems are also discussed.

## 1. Introducción

La evaluación educativa y, en concreto, la evaluación externa de las competencias de los alumnos, ha pasado en menos de quince años a ocupar un lugar central entre los instrumentos de la política educativa. La evaluación de competencias, impulsada en buena medida por instituciones globalizadoras como la OCDE, forma parte actualmente de la agenda de reformas educativas muy diversas. Desde aquellas orientadas a la intensificación de los cuasimercados educativos y la elección, hasta las que ponen un mayor acento en la producción pública y en la rendición de cuentas de los centros públicos, las agendas reformistas otorgan a la evaluación el carácter de piedra angular de los diferentes diseños.

El desarrollo de la evaluación de competencias refleja, en todo caso, un énfasis en los *resultados* del sistema educativo y, en concreto, en el logro de los alumnos. No se trata de algo trivial, sino de un desplazamiento en toda regla del foco de interés en los análisis e intervenciones sobre el sistema educativo, desde los *inputs* y procesos hacia las competencias adquiridas. En el ámbito de la equidad, por ejemplo, la evaluación de competencias está en sintonía con la selección de la igualdad de logros como el objetivo prioritario de los sistemas (véase Baye *et al.*, 2006), en contraposición a la igualdad de acceso o la igualdad de trato. En el ámbito de la eficiencia, resulta evidente que la generación y disponibilidad de datos relativos a uno de los «productos finales» más importantes del sistema resulta de gran interés.

Podríamos, pues, hablar de una «colonización» de los sistemas educativos por parte de la evaluación de competencias. En aquellos países en los que esta colonización ha sido más intensa, como los Estados Unidos, se han establecido fuertes vínculos entre financiación de los centros y resultados de la evaluación y se han hecho más notorios algunos de los efectos no deseables de la evaluación educativa (véase Ravitch, 2010). El balance entre los beneficios y costes de la evaluación educativa no puede tomarse, de forma acrítica, como positivo *per se*: como veremos, en muchos contextos reales la evaluación genera problemas de eficiencia y/o equidad que merecen una atención cuidadosa por parte de los diseñadores y aplicadores de las políticas educativas.

Parte del éxito innegable de la propagación de la evaluación educativa ha estado basado en el impacto mediático de los programas internacionales de evaluación de competencias y, más en concreto, del programa PISA de la OCDE. La posibilidad de una comparación inmediata (y a menudo poco reflexiva) de los resultados educativos entre países y, en España, entre comunidades autónomas, ha capturado la atención de los medios de comunicación y de la opinión pública.

Teniendo en cuenta el entorno relativo a las políticas educativas que hemos comentado, en este artículo nos planteamos como objetivo describir y valorar los efectos de la evaluación de competencias, prestando especial atención a las aplicaciones en entornos empíricos reales y, más en concreto, a las aplicaciones en España. Para ello, en el apartado 2 se sistematiza una serie de papeles alternativos que se otorga a la evaluación de competencias desde diferentes aproximaciones a la política educativa, con objeto de arrojar luz sobre la naturaleza compleja de la evaluación. Parte de esta naturaleza se plasma sobre el papel que se otorga a la difusión de resultados y, en concreto, la publicación de rankings de centros. El apartado 3 está destinado a la descripción del «mapa» actual de los programas de evaluación de competencias llevados a cabo por las instituciones internacionales y nacionales. En el apartado 4 discutimos y valoramos diversos efectos (algunos de ellos deseados, otros no deseados) que genera la introducción de la evaluación de competencias en los sistemas educativos.

## 2. Un único instrumento, varias finalidades

La evaluación es considerada, desde diferentes aproximaciones de política educativa y por motivos alternativos, como necesaria y muy importante. Se podría llegar a estimar, incluso, como un término polisémico, si no fuera porque en realidad el instrumento descrito

como necesario es el mismo en todas las aproximaciones, variando sus funciones y, sobre todo, la difusión de los resultados. Intentaremos sintetizar en los próximos párrafos la ubicación de la evaluación en tales alternativas.

a) *Evaluación para la elección.* Desde una aproximación estrictamente centrada en la capacidad de elección, y enmarcada en una política pública cuyo objetivo consista en el funcionamiento del sistema del modo más próximo a un mercado, la evaluación es un instrumento crucial, en tanto que proporciona información para una elección eficaz por parte de las familias. En este sentido, los diseños de cuasimercados educativos (véase Glennerster, 1991) enfatizan la necesidad de hacer llegar la información a las familias.

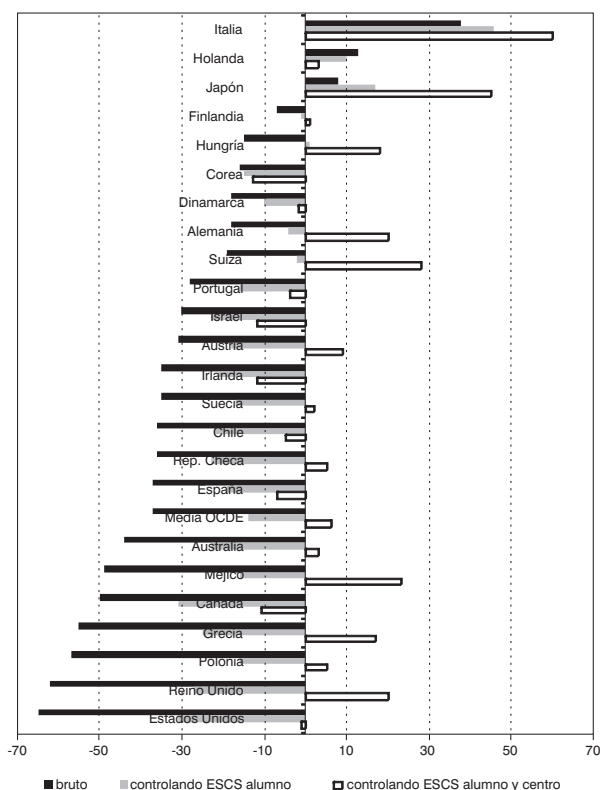
b) *Evaluación para la rendición de cuentas de los centros públicos.* La combinación de autonomía de los centros más evaluación de resultados está orientada a garantizar que las decisiones autónomas de los centros educativos tengan como contrapartida la rendición de cuentas de éstos al financiador. En el extremo de este diseño, la responsabilidad del sector público se podría limitar a la evaluación, lo que nos lleva al concepto de «estado evaluador»<sup>1</sup>. En esta aproximación, a diferencia de la anterior, la difusión de los resultados de la evaluación entre los potenciales usuarios no es imprescindible. De hecho, resulta coherente con esta aproximación la limitación o prohibición de la difusión de los resultados: ésta puede reforzar los procesos de segregación y generar problemas adicionales de equidad. Los centros con mejores puntuaciones en la evaluación pueden atraer más demanda «selecta», es decir, de aquellas familias con más recursos, con mayor capacidad a la hora de elegir y hacer efectiva su elección de centro; por contra, los centros con peores puntuaciones ven empeorada su situación debido a la «huida» de los pocos alumnos con buenos recursos familiares, que se desplazan hacia centros de mayor puntuación. La combinación de estos procesos con los «efectos compañero» acentúa, además, el resultado de segregación adicional. Este tipo de reflexión es la que ha llevado al legislador español a limitar la difusión de los resultados de las evaluaciones establecidas en la Ley Orgánica de Educación de 2006<sup>2</sup> y en la Ley de Educación de Cataluña de 2009<sup>3</sup>.

El debate de política educativa en torno a la conveniencia o no de distribuir información pública de los resultados de las evaluaciones en forma de rankings nos lleva necesariamente a preguntarnos sobre el formato posible de estos rankings y sus implicaciones. En concreto, quisiéramos distinguir entre los rankings de resultados «brutos» y los rankings de valor añadido contextual. Los primeros se corresponden a la puntuación media directamente extraída de los resultados de la evaluación de las competencias. Dependen, por tanto, en muy buena medida, de la composición socio-cultural y económica de los alumnos del centro e incorporan, además, los efectos de la interacción entre ellos en el aula (*peer-effects*). Los rankings que se publican en España<sup>4</sup> son de este tipo: en ellos, por ejemplo, una posición muy favorable de un centro puede ser atribuible, indistintamente, a la calidad que consigue incorporar a sus procesos el centro (con independencia del tipo de usuarios) y al origen de sus alumnos y las interacciones que se producen entre ellos. Por el contrario, en los rankings de valor añadido contextual (véase Ray *et al.*, 2009) en las puntuaciones de los centros se controla el efecto del tipo de alumnado, de tal modo que la puntuación recoge exclusivamente lo que aporta el centro por sí mismo. Este control se puede llevar a cabo mediante modelos de regresión aunque, en el caso del Reino Unido, por ejemplo, se les explica a las familias recurriendo al argumento de que se compara el centro con aquellos que tienen una composición similar de sus usuarios.

Puede observarse en el gráfico 1 un ejemplo de la diferencia entre las puntuaciones «en bruto» y las puntuaciones de valor añadido contextual en el caso de la evaluación de competencias de lectura de PISA-2009. En este caso, se presentan las diferencias existentes entre los centros públicos y los privados con tres mediciones: la «bruta»; la de valor añadido en la que se ha descontado el efecto del origen socio-cultural y económico de los alumnos (medido con el índice ESCS de PISA) y la de valor añadido en la que se ha descontado el efecto anterior y, adicionalmente, el mismo efecto medido de forma agregada para el conjunto de alumnos del centro. Puede verse cómo, para la mayoría de los países evaluados, las diferencias claramente a favor de los centros privados cuando la medición es «en bruto» pasan a ser considerablemente menores, inexistentes o incluso a favor de los centros públicos cuando la medición es de valor añadido contextual.

**Gráfico 1**

**PISA-2009. Diferencias en la puntuación de competencias de lectura entre centros públicos y centros privados. Diferencias brutas y descontando los efectos del estatus económico y socio-cultural (índice ESCS) del alumno y los efectos combinados del ESCS del alumno y el centro. Selección de países de la OCDE**



*Fuente:* elaborado a partir de la tabla 3.6 (Anexo 3) de Instituto de Evaluación (2010) y de OECD (2010).

*Nota:* los valores negativos indican diferencias de puntuación a favor de los centros privados; los valores positivos diferencias a favor de los centros públicos.

En una variante de los rankings de valor añadido, en lugar de comparar a cada centro con los «similares», se compara al centro con sí mismo en un momento anterior del tiempo. Se trata, en este tipo de ranking, de valorar la evolución a lo largo del tiempo en función de cambios aplicados (en los recursos o en la dirección, por ejemplo), valoración que tiene más interés para las administraciones educativas que para los usuarios.

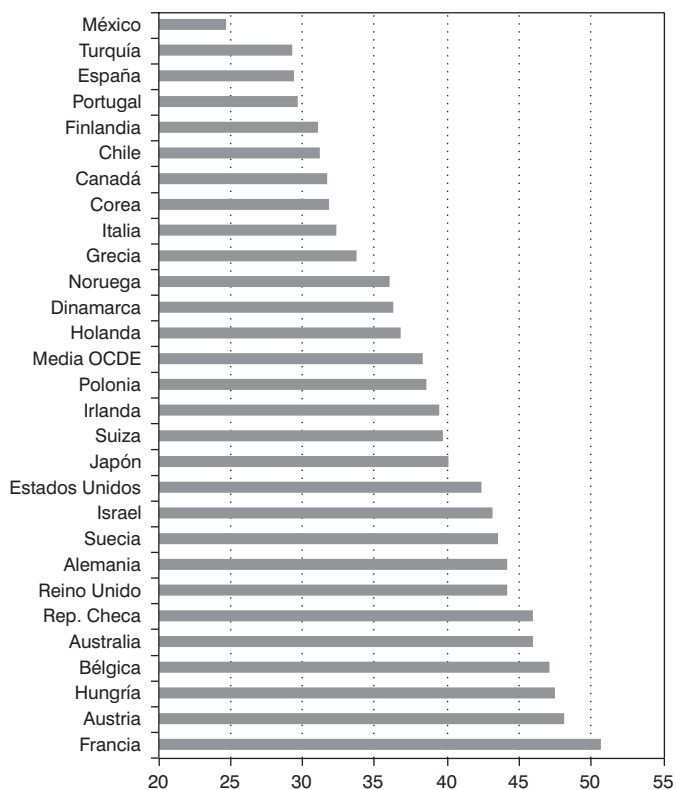
Podríamos cuestionar qué tipo de ranking puede ser de interés para cada agente. Una familia únicamente interesada en las puntuaciones de sus hijos en las evaluaciones de competencias, sin altruismo, posiblemente tenga interés en los rankings de resultados «en bruto», que incorporan el efecto de la interacción con los compañeros potenciales de sus hijos. La administración educativa, así como algún tipo de usuario con criterio de decisión diferente al descrito anteriormente, estará más interesada en los rankings de valor añadido. Resulta claro que los rankings relevantes para la administración, en el caso de desear vincular la financiación con los resultados de la evaluación, son los de valor añadido, especialmente si tenemos en cuenta que el stock de usuarios con un origen social favorable para los resultados de las evaluaciones es limitado y no ampliable en el corto y medio plazo.

c) *Evaluación destinada a la información interna de los centros.* En el marco de un funcionamiento relativamente autónomo de los centros, otro tipo de evaluación que puede considerarse es la destinada específicamente a una valoración interna por parte del centro, con objeto de mejorar los procesos. La definición de las evaluaciones de diagnóstico en la LOE (art. 144.2) apunta en este sentido: «En el marco de sus respectivas competencias, corresponde a las Administraciones educativas desarrollar y controlar las evaluaciones de diagnóstico en las que participen los centros de ellas dependientes y proporcionar los modelos y apoyos pertinentes a fin de que todos los centros puedan realizar de modo adecuado estas evaluaciones, que tendrán carácter formativo e interno».

d) *Evaluación orientada a obtener información general del funcionamiento del sistema y a la investigación educativa.* Los tipos de evaluación a los que nos hemos referido en a), b) y c) exigen una recogida censal (en el nivel de los centros, no necesariamente de los alumnos) de información. Por contra, en este último tipo la recogida de información puede ser muestral, al no ser necesario contar con información de cada centro sino únicamente la que proviene de una muestra suficientemente representativa. En realidad, las evaluaciones de competencias pioneras han sido usualmente muestrales y únicamente en los últimos años se han extendido algunas de ellas a la totalidad de los centros.

La información que se obtiene por medio de las evaluaciones «de sistema» puede ser utilizada directamente por las administraciones educativas o bien ser elaborada en investigaciones cuyos resultados pueden, eventualmente, ser transferidos a las administraciones educativas. En esta segunda vía resulta muy llamativo, por ejemplo, el despliegue de investigación, en España, que analiza los determinantes de los resultados educativos por medio de las evaluaciones de PISA, específicamente a partir de la segunda mitad de la década de 2000 (puede consultarse un ejemplo concreto en Calero y Escardíbul, 2007 y una revisión general de este tipo de estudios en Cordero *et al.*, 2013). Quizás uno de los aspectos que esta investigación ha permitido establecer con mayor precisión ha sido el conjunto de relaciones que se establecen entre el origen sociocultural y económico de los alumnos y sus resultados en la evaluación de competencias. Aportamos aquí, en los gráficos 2 (para diversos países de la

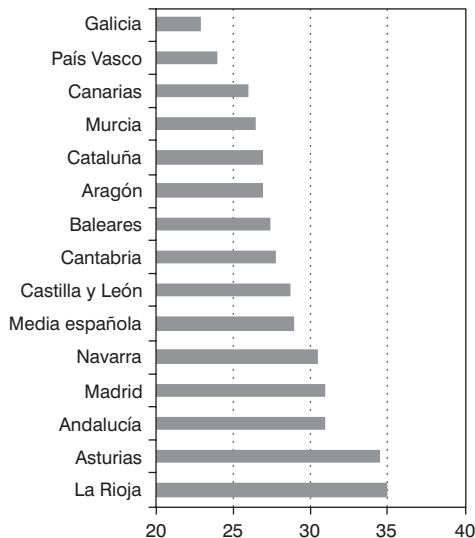
**Gráfico 2**  
**Puntos de variación en la competencia de lectura por cada punto que varía el índice ESCS.**  
**Selección de países de la OCDE**



*Fuente:* elaborado a partir de OECD (2010).

OCDE) y 3 (para las comunidades autónomas con muestra representativa en PISA-2009) los resultados de un indicador que sintetiza esta relación. Se trata de la pendiente de la recta de regresión entre el índice ESCS y la puntuación en la competencia de lectura. Esta pendiente indica cómo transforma el sistema educativo las desigualdades previas (reflejadas en el índice ESCS) en desigualdades de logro educativo y es, por tanto, una excelente síntesis de la capacidad del sistema educativo para atemperar o, por el contrario, reproducir de forma ampliada las desigualdades sociales previas.

**Gráfico 3**  
**Puntos de variación en la competencia de lectura por cada punto que varía el índice ESCS.**  
**Comunidades Autónomas y media española.**



*Fuente:* elaborado a partir de OECD (2010).

### 3. Un «mapa» de la evaluación de competencias realizada por organismos internacionales y nacionales

El organismo pionero en materia de evaluación de competencias fue la IEA (International Association for the Evaluation of Educational Achievement), fundada en 1958, que ha desarrollado más de treinta evaluaciones comparativas internacionales. Las más influyentes, por sí mismas y por su incidencia metodológica sobre PISA, han sido TIMSS (Trends in International Mathematics and Science Study), con una periodicidad cuatrienal desde 1995, y PIRLS (Progress in International Reading Literacy Study), con una periodicidad quinquenal desde 2001. Se trata de estudios basados en muestras; en España se ha aplicado TIMSS en las ediciones de 2003 y 2011 y PIRLS en las ediciones de 2006 y 2011.

La IEA, adicionalmente, efectúa evaluaciones internacionales sobre otras competencias, como las de civismo y ciudadanía en ICCS-2009 (International Civic and Citizenship Education Study), o las relativas a las TICs en ICILS-2013 (International Computer and Information Literacy Study). En estas dos evaluaciones participó España con una muestra representativa<sup>5</sup>. Dos programas pioneros que está llevando a cabo la IEA, en los que también participa España, son el PPP (Preprimary Project), estudio longitudinal sobre la calidad de vida de los niños en edad de asistir a la educación infantil y, desde 2005, TEDS-M (Teacher Education and Development Study in Mathematics), donde se estudia la formación del profesorado de matemáticas de primaria y secundaria.

El interés de la OCDE en la evaluación de competencias se ha plasmado, desde el año 2000, en una diversidad de evaluaciones. Obviamente, ha sido el programa PISA el que ha tenido un mayor impacto, haciendo pasar los resultados educativos a las portadas de los periódicos y, también, estableciendo unos procesos metodológicos y técnicos que se han convertido en hegemónicos en la evaluación de competencias. Se trata, como dijimos en la introducción, de un importante esfuerzo «globalizador» en las políticas educativas: piénsese, por ejemplo, en la voluntad de medir competencias idénticas en 65 países con culturas y sistemas educativos radicalmente diversos.

Los tres tipos de competencias seleccionadas en PISA (lectura, matemáticas y ciencias) no son las únicas en las que la OCDE está interesada. Ya en la edición de 2009 de PISA se introdujo, en una muestra reducida de 19 países, una evaluación de competencias relativas a las habilidades para adquirir y evaluar información en Internet (Digital Literacy Assessment, véase OECD, 2011). Resulta también destacable la introducción, en PISA-2012, de la evaluación de competencias financieras (PISA Financial Literacy Study 2012).

Otro ámbito muy relevante de evaluación de competencias que está desarrollando la OCDE es el correspondiente a las competencias cognitivas de la población adulta, de 16 a 65 años, en el PIAAC (Programme for the International Assessment of Adult Competencies), aplicado de forma cíclica a partir del año 2011. Finalmente, la OCDE tiene un programa (TALIS —Teaching and Learning International Survey—) de estudio del profesorado, si bien éste no se orienta a la evaluación de competencias.

La Comisión Europea se ha unido a las instituciones internacionales que realizan evaluación de competencias. En 2011 se aplicaron las pruebas, por primera vez, del Indicador Europeo de Competencia Lingüística, orientado a proporcionar información sobre la capacidad de comprensión oral, lectora y la escritura de las dos lenguas europeas más enseñadas en cada país. La evaluación se lleva a cabo mediante una muestra representativa en el nivel estatal y se aplica a los estudiantes del último curso del primer ciclo de enseñanza secundaria (cuarto de ESO en España). Esta iniciativa se enmarca en una decisión del Consejo Europeo de Barcelona (2002) recogida, posteriormente, en la estrategia Education and Training 2020, que fijó en 2009 como uno de sus objetivos el aprendizaje de dos idiomas extranjeros desde edad temprana. La falta de información homogénea sobre el cumplimiento de tal objetivo motivó el desarrollo del programa al que nos referimos.

En España, la aprobación de la Ley Orgánica de Educación de 2006 supuso la introducción a gran escala de la evaluación de competencias. Los dos instrumentos que se introdujeron fueron los siguientes:

- Evaluaciones generales de diagnóstico. Se trata de evaluaciones muestrales, orientadas a obtener información general del funcionamiento del sistema. Se han aplicado en cuarto curso de Primaria y en segundo curso de ESO<sup>6</sup>. Las competencias evaluadas en ambos casos son las de comunicación lingüística, matemáticas, conocimiento e interacción con el mundo físico y competencias sociales y ciudadanas.
- Evaluaciones de diagnóstico. Son evaluaciones censales que se aplican en cuarto de Primaria y segundo de ESO. Su diseño y aplicación es responsabilidad de las administraciones educativas de las comunidades autónomas. Están orientadas a propor-



cionar información interna a los centros y pueden ser utilizadas por las administraciones educativas para valorar el funcionamiento de centros concretos.

Algunas comunidades autónomas han optado por desarrollar instrumentos adicionales de evaluación. En Cataluña, por ejemplo, se aplica desde 2009 una evaluación censal en sexto de Primaria y, desde 2012, una evaluación censal en cuarto de ESO.

Finalizaremos este apartado haciendo referencia a las instituciones evaluadoras en España. Además del Instituto Nacional de Evaluación Educativa, dependiente del Ministerio de Educación, Cultura y Deporte, las comunidades autónomas han establecido unidades de evaluación. Podríamos establecer una tipología compuesta por tres tipos de unidades de evaluación. En un primer grupo estarían aquellas dependientes directamente de la consejería de educación correspondiente, integradas en el organigrama como una dirección o subdirección general. A un segundo grupo de unidades se les ha concedido una autonomía relativa dentro de la consejería correspondiente; es el caso, por ejemplo, del Instituto Vasco de Evaluación e Investigación Educativa, el Instituto Valenciano de Evaluación y Calidad Educativa o el Consejo Superior de Evaluación del Sistema Educativo de Cataluña. El tercer grupo de unidades son agencias como la Agencia Andaluza de Evaluación Educativa, creada en 2008, o la Agencia de Evaluación y Prospectiva de la Educación, de Cataluña, creada en 2009 pero cuyo inicio de actividades fue pospuesto *sine die* en 2011. La creación de estas agencias tiene como objetivo proporcionar una mayor autonomía a la actividad evaluadora y evitar, por tanto, la poca deseable coincidencia de evaluador y evaluado. Parece deseable que en los próximos años se avance en el establecimiento de unidades de evaluación cada vez más autónomas, que permitan aproximar nuestros sistemas de evaluación educativa a los más avanzados en el contexto internacional.

#### **4. Algunos efectos (deseados y no deseados) de la evaluación de competencias**

Es preciso tener en cuenta que en una diversidad de ámbitos de las ciencias sociales (y naturales, por supuesto) la mera observación altera el comportamiento de lo observado. En la evaluación de competencias se cumple esta afirmación; uno de los ámbitos en los que más claro resulta es su incidencia sobre el currículum educativo. El cambio de enfoque, en los últimos años, en la organización y definición de los contenidos en la programación curricular, la adaptación a una programación según competencias, ha sido paralelo a la implantación de sistemas de evaluación externos, precisamente centrados en las competencias. En nuestro sistema educativo, por ejemplo, esta adaptación ha tendido a arrinconar las prácticas basadas en el aprendizaje memorístico. El concepto de competencia es instrumental; para exponerlo utilizaremos, entre las diferentes alternativas disponibles, la definición aportada por la OCDE (OECD, 2005: 4): «Una competencia es algo más que únicamente conocimientos y habilidades. Implica la habilidad para enfrentarse a demandas complejas, haciendo uso y movilizandolos recursos psicosociales (que incluyen habilidades y actitudes) en un contexto particular. Por ejemplo, la habilidad para comunicarse de forma efectiva es una competencia que puede hacer uso del conocimiento del lenguaje de un individuo, sus habilidades prácti-

cas de las tecnologías de la información y de actitudes hacia aquéllos con quien se comunica la persona».

La selección de las competencias consideradas como básicas y merecedoras, por tanto, de ser evaluadas, no es trivial. La selección, por ejemplo, de las competencias básicas en materia de lectura, matemáticas y ciencias y, adicionalmente, como hemos mencionado anteriormente, de competencias financieras y de adquisición de información en Internet por parte de la OCDE en el programa PISA, ha trasladado hacia los sistemas educativos (directa e indirectamente) el énfasis en esas competencias y no en otras. Se podrían discutir, por ejemplo, los criterios que utiliza la OCDE para seleccionar para su evaluación internacional esas competencias y no otras relevantes y transmitidas por el sistema educativo, como las vinculadas con *outputs* no cognitivos (entre los que figuran valores como la justicia, la igualdad, la democracia, la responsabilidad y la tolerancia)<sup>7</sup>.

Los efectos de la evaluación educativa no se circunscriben al ámbito de lo deseado en una visión «ingenua» de las posibilidades de mejora del sistema. La incidencia de la evaluación sobre los contenidos puede conducir a que «sólo se aprenda lo que se evalúa». En el extremo, puede llegarse a una situación en la que los contenidos de las materias se estrechan y contorsionan de tal modo que se adecúan exclusivamente a superar los tests de competencias («*teaching by the test*»). Para utilizar una expresión bien conocida, la evaluación deja de ser un medio para convertirse en un fin. Este riesgo, sin embargo, puede controlarse mediante una definición suficientemente amplia de las competencias evaluadas. Refiriéndonos a la competencia de comprensión lectora, si lo que se evalúa es, por ejemplo, la comprensión de un artículo de prensa, no parece que la adecuación de la docencia a este tipo de comprensión implique pérdidas o reducciones inaceptables de la diversidad o calidad de la transmisión de conocimientos.

Otro conjunto de efectos, estos sí clasificables claramente en la zona de los no deseados, tienen que ver con lo que podríamos denominar efecto de «cama de Procasto». Especialmente en entornos en los que la financiación se vincula a la evaluación, los centros tienen incentivos para focalizar los esfuerzos de la docencia en aquellos alumnos «recuperables», prestando una menor atención tanto a los alumnos situados muy por debajo de los umbrales marcados por las competencias mínimas establecidas como a aquellos que se sitúan suficientemente por encima de ellos. Estos comportamientos estratégicos de los centros<sup>8</sup> suponen un claro empobrecimiento tanto en términos de excelencia como en términos de las posibilidades de las políticas compensatorias para los alumnos con competencias más bajas.

Nos referiremos, finalmente, a los efectos generales de las evaluaciones internacionales en el nivel político. En el lado positivo, estas han despertado el interés por la situación de otros países, bajo la convicción de que el sistema educativo español puede aprender de otras experiencias exitosas. Así, ya no resultan tan extrañas las alusiones a sistemas educativos considerados de referencia, como los de países del Este asiático o Finlandia.

Ahora bien, ello también tiene una cara negativa, ya que las evaluaciones internacionales son utilizadas en ocasiones de forma poco adecuada, siendo interpretadas como meros rankings o competiciones entre países, sin tenerse en cuenta sus trayectorias y características propias. A modo de ejemplo, suele hablarse del éxito de los países asiáticos en TIMSS y

PISA, sin considerarse una cuestión tan fundamental como su tamaño: se contabilizan por igual, a efectos de ranking entre naciones, a países como Japón o Corea, a antiguas ciudades-estado como Hong Kong o, incluso, a ciudades como Shanghai. La utilización de las evaluaciones internacionales como rankings también tiende a obviar la heterogeneidad en las poblaciones y las propias limitaciones metodológicas de las evaluaciones por lo que, en definitiva, puede conducir a la extracción de conclusiones imprecisas y a la aplicación de políticas erróneas.

## 5. Conclusiones

La evaluación de competencias se ha constituido, en los últimos años, como una de las piedras angulares de las políticas educativas. Esta afirmación es válida en prácticamente todos los contextos, debido a la diversidad de funciones alternativas que se han atribuido a los sistemas de evaluación. Hemos visto, en este artículo, cómo la evaluación se utiliza como elemento básico de las políticas orientadas a la elección, pero también de políticas orientadas al refuerzo de la rendición de cuentas de los centros públicos, entre otras. La evaluación ha «colonizado», en buena media, los sistemas educativos, en un movimiento notable vinculado con los procesos de globalización educativa.

Un aspecto crucial de los diseños de la evaluación educativa es la decisión en torno a la publicación de resultados en forma de rankings de centros. Esta decisión es dependiente del tipo de política educativa en la que esté inserto el sistema de evaluación y oscila entre la publicación de rankings de puntuaciones «en bruto» y la restricción total de publicación de rankings, pasando por la publicación de rankings de valor añadido.

En el artículo hemos revisado el importantísimo papel de instituciones internacionales como la IEA y la OCDE en el impulso y la difusión de la evaluación de competencias. El programa PISA es paradigmático en este sentido. Se ha constituido en hegemónico en cuanto a definición de competencias, metodologías y técnicas; simultáneamente, ha capturado la atención de los medios de comunicación y la opinión pública.

En España, las diversas evaluaciones que se establecen a partir de la Ley Orgánica de Educación de 2006 se enmarcan en la tendencia señalada por la IEA y la OCDE, con diversos matices (como la inclusión de las competencias sociales y ciudadanas). Las unidades de evaluación educativa creadas, desde la década de 1990, en España, se han diseñado estableciendo un nivel de autonomía usualmente nulo o reducido con respecto a las consejerías de educación correspondientes. Únicamente la más reciente creación de agencias apunta en el sentido de proporcionar suficiente distancia entre evaluador y evaluado y permite el acercamiento eventual a los sistemas evaluativos más avanzados en el contexto internacional.

La aplicación generalizada de los sistemas de evaluación de competencias está teniendo, sin duda, algunas consecuencias positivas, entre las que podríamos destacar la reorientación de la planificación curricular hacia las competencias. Sin embargo, no debemos olvidar los potenciales efectos no deseados que genera la implementación de sistemas de evaluación de competencias. Entre ellos hemos destacado el estrechamiento y contorsión de los conte-

nidos transmitidos («teaching by the test») y el efecto de «cama de Procusto», que tiende a focalizar la docencia en aquellos alumnos susceptibles de elevar su puntuación por encima del umbral fijado por la evaluación.

En definitiva, la evaluación de competencias es un instrumento muy potente para la política educativa y sus muy diversos diseños pueden adaptarse a objetivos muy diferentes. No resulta conveniente adoptar una actitud ingenua con respecto a sus efectos: la multiplicidad de aplicaciones empíricas en los últimos quince años nos proporcionan ya una información consistente en torno a las luces y sombras de la evaluación educativa.

## Notas

1. Sobre el concepto de «estado evaluador», véase Elliot (2002).
2. El redactado del artículo 140.2 de la LOU es el siguiente: «La finalidad establecida en el apartado anterior no podrá amparar que los resultados de las evaluaciones del sistema educativo, independientemente del ámbito territorial estatal o autonómico en el que se apliquen, puedan ser utilizados para valoraciones individuales de los alumnos o para establecer clasificaciones de los centros».
3. El redactado del artículo 184, c) de la Ley de Educación de Cataluña establece que la evaluación educativa tiene que estar sometida al principio de «uso reservado de la información individualizada de los agentes y de los centros y servicios educativos respecto a la evaluación general del sistema».
4. Es el caso de la publicación, por parte de la Comunidad Autónoma de Madrid, de los resultados de la Prueba de Conocimientos y Destrezas Imprescindibles en sexto de Primaria.
5. Cataluña había participado en un estudio previo relativo al uso de las TICs en la docencia de matemáticas y ciencias, SITES-2006 (Second Information Technology in Education Study).
6. Sin embargo, en el *Plan de Acción 2010-2011*, aprobado en el Consejo de Ministros de 25 de junio de 2010, se planteó una modificación que supondría su aplicación como evaluaciones de final de etapa, en sexto de Primaria y cuarto de ESO.
7. Para una aproximación a la evaluación de *outputs* educativos no cognitivos, véase García-Valiñas *et al.* (2012).
8. Véase Toch (2006) para una revisión de estos comportamientos en Estados Unidos a partir de la aplicación de la ley *No Child Left Behind* de 2003.

## Bibliografía

- Baye, A. *et al.* (2006): *A Set of Indicators to Measure Equity in 25 European Union Education Systems. Report for the European Commission*. Liège: Service de Pédagogie.
- Calero, Jorge y Josep-Oriol Escardíbul (2007): «Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003». *Hacienda Pública Española*, n. 83 (4/2007), pp. 33-66.
- Cordero, J. M., E. Crespo y F. Pedraja (2013): «Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España». *Revista de Educación*, DOI: 10.4438/1988-592X-RE-2011-362-161.

- Elliot, John (2002): «The paradox of educational reform in the evaluatory state: implications for teacher education». *Prospects*, vol. XXXII, n. 3.
- Gairín, J. (2009): «Usos y abusos de la evaluación. La evaluación como autorregulación», en Gairín, J. (ed.) *Nuevas funciones de la evaluación*. Madrid, Ministerio de Educación.
- García-Valiñas, M., M. Muñoz-Pérez y J. Suárez-Pandiello (2012): «The determinants of non-cognitive education: does school matter? Empirical evidence from Spain». *XIX Encuentro de Economía Pública*. Santiago de Compostela.
- Glennerster, H. (1991) «Quasi-Markets for Education?» *The Economic Journal*, vol. 101, n. 408, pp. 1268-1276.
- Instituto de Evaluación (2010): *PISA. Programa para la evaluación internacional de los alumnos OCDE. Informe español*. Madrid: Instituto de Evaluación.
- OECD (2011): *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI)*. París: OCDE.
- OECD (2010): *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*. París: OECD.
- OECD (2005): *The definition and selection of key competencies*. París: OECD.
- Ravitch, Diane (2010): *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.
- Ray, Andrew, Helen Evans y Tanya McCormack (2009). «School Value Added Measures in England». *Revista de Educación*, n. 348, pp. 47-66.
- Toch, Thomas (2006): *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*. Education Sector Reports.