

Las evaluaciones y pruebas de diagnóstico como instrumentos de la política educativa: significado y balance de su aplicación

F. Javier Merchán Iglesias
Fedecaria-Sevilla

La medición del rendimiento escolar

En los últimos años, en muchos países y a nivel internacional, al margen de las calificaciones de las evaluaciones curriculares, ha proliferado la realización de pruebas estandarizadas para medir el grado de aprendizaje que adquieren los alumnos en los centros escolares. En cierto sentido no se trata de una estricta novedad. En distintos sistemas educativos, hace tiempo que los alumnos son sometidos a pruebas comunes, cuyos resultados determinan el ingreso en establecimientos superiores o el paso de un ciclo de enseñanza a otro. Estas pruebas tienen un carácter selectivo, es decir, su función es la de determinar quiénes pueden acceder a un cierto nivel de la enseñanza o la de acreditar una titulación. Las pruebas que tienen como objetivo la inspección del sistema o la recogida de información sobre el funcionamiento de la educación tampoco son nuevas. En los primeros años del siglo pasado, los planteamientos conductistas, de los que Edwards L. Thorndike fue uno de sus principales teóricos y valedores, promovieron la realización de mediciones sobre los efectos de la educación: "Thorndike creía firmemente en el valor del experimento y la medición. «Todo lo que existe», decía, «existe en cierta cantidad. Conocerlo exhaustivamente implica conocer su cantidad lo mismo que su calidad»" (Ebel, 1993, p. 92). En 1903 Thorndike publicó el primer libro de texto sobre medición educacional. Sus pasos fueron seguidos por Binet, que

en 1904, a petición del Ministerio de Educación francés, elaboró pruebas para medir la inteligencia de los alumnos. En 1928, Ben D. Wood y Williams S. Learned realizaron en Pennsylvania pruebas para medir el rendimiento educativo, al igual que ocurrió en 1929 en el estado de Iowa. Ya en 1947 se estableció en los Estados Unidos de América el *Educational Testing Service*, que tenía como misión la comprobación de los niveles de los alumnos que aspiraban a la admisión en el *College*, y poco después, en 1956, Tyler y Bloom establecieron una taxonomía jerárquica de objetivos educativos mensurables.

Aunque todas estas experiencias e iniciativas no estuvieron exentas de críticas –discutiéndose, sobre todo, que la consecución de resultados y objetivos operativos debiera ser la meta principal de la educación–, desde los años cincuenta del pasado siglo van apareciendo distintas actuaciones para poner en marcha programas internacionales de evaluación cuyos objetivos ya nada, o muy poco, tienen que ver con las posibilidades de promoción de los estudiantes o la evaluación de sus capacidades. En el contexto del movimiento de la eficacia escolar y de la mejora de la educación, así como de los procesos de reforma, surge en 1958 la *International Association for the Evaluation of Educating Achievement* (IEA) en torno al Instituto de Educación de la UNESCO¹. Desde entonces esta institución ha venido realizando estudios en diversas áreas, de entre los que destacan los programas TIMSS (*Trends in International Mathematics and Science Stu-*

¹ Vid. <http://www.iea.nl/>

dy), PIRLS (*Progress in International Reading Literacy*) y el ICCS (*International Civic and Citizenship Education Study*).

Pero no cabe duda de que, de entre todos los estudios sobre el rendimiento de los alumnos, el que ha adquirido mayor notoriedad en los últimos años es el que se conoce como PISA (*Programme for International Student Assessment*) y que está dirigido por la OCDE. Se trata de un programa iniciado a finales de los años 90, mediante el cual, a partir del año 2000, se vienen realizando pruebas a alumnos de 15 años². En comparación con otros estudios, el PISA aporta novedades significativas tanto desde el punto de vista de su confección como del sentido que tiene y se le atribuye a los resultados y a su propia realización. Las pruebas PISA no se refieren a contenidos del currículum³ sino a las *competencias básicas*, es decir, a la capacidad que tienen los estudiantes de extrapolar lo que han aprendido y aplicar sus conocimientos en nuevas circunstancias, a su relevancia para el aprendizaje a lo largo de la vida y a su regularidad. El *enfoque de competencias*, que ha sido puesto de actualidad desde el mundo de la empresa y difundido por la OCDE, implica que se evalúa cómo los alumnos pueden hacer uso de su capacidad lectora para comprender distintos tipos de material escrito con el que probablemente se van a encontrar al gestionar su vida diaria; de qué forma pueden utilizar su competencia matemática para resolver distintos tipos de retos y problemas relacionados con las matemáticas; y el modo en que los alumnos puedan hacer uso de sus conocimientos y destrezas científicas para comprender e interpretar distintos contextos científicos (MEC, 2007, p. 16).

Por otra parte, la información que recoge PISA no se limita a medir el nivel de competencias adquiridas por los alumnos, sino que se extiende a otros aspectos que hipotéticamente inciden en su rendimiento. Así, PISA 2006 preguntó a los alumnos sobre el

entorno familiar, su opinión sobre distintos aspectos relacionados con las ciencias, sobre su actitud hacia el medio ambiente, sus expectativas profesionales y su vinculación con las ciencias, el tiempo dedicado al estudio, las clases de ciencias en el colegio y el uso de ordenadores. Los directores de los centros escolares respondieron a preguntas sobre las características del centro, el alumnado, los recursos, el profesorado, la organización, la inclusión del medio ambiente en el currículo y la orientación profesional y la preparación de los alumnos para estudios posteriores.

Los objetivos explícitos de las evaluaciones y pruebas de diagnóstico

Como se ha dicho, el objetivo de este tipo de pruebas va más allá de la mera recogida de información sobre los resultados que obtienen los alumnos. El uso del término diagnóstico pone de manifiesto la concomitancia que se pretende con los procesos que se utilizan en el campo de la medicina: la evaluación, a partir de los datos que suministran las pruebas, facilitaría información sobre los problemas, actuando como diagnóstico, tras el cual se procede a poner en marcha *medidas de curación*, es decir, propuestas de mejora, tanto al nivel de cada centro, mediante los planes anuales y la autoevaluación, como del conjunto del sistema, mediante la política educativa. Según este planteamiento, las pruebas o evaluaciones de diagnóstico forman parte de un dispositivo de mejora de la educación, siendo su función principal la de establecer el punto de partida de donde se derivan las actuaciones para mejorar, ya que producen un conocimiento capaz de explicar las dificultades que impiden alcanzar los resultados deseables, y sugieren líneas de actuación para lograrlo. En este sentido nos encontramos ante una reconceptualización

² PISA 2000 estuvo destinado a hacer una evaluación más profunda en lectura, PISA 2003 se centró en matemáticas, y en la edición de 2006 le ha tocado el turno a las ciencias.

³ Entre otras razones porque de ser así hubiera sido imposible hacer comparaciones internacionales de los resultados, ya que a veces existen diferencias significativas en el currículum de los países.

del concepto de evaluación, que, frente al modelo de procesos, se plantea ahora como análisis de resultados.

Así, PISA no es meramente un estudio que se limite a describir el rendimiento de los alumnos ante las pruebas, sino que –y ésta es otra de sus diferencias respecto a otras pruebas– trata también de ofrecer explicaciones de los resultados. A este respecto la metodología del estudio contempla el trabajo con una serie de variables explicativas: a) Dotaciones iniciales; b) Variables de esfuerzo individual; y c) Variables del centro escolar⁴. En lo que se refiere a las dotaciones iniciales, se considera el nivel educativo de los padres, su ocupación, nivel económico o de riqueza, posesiones educativas, nacionalidad, sexo, lengua en la que normalmente se habla... A tenor de la contrastada influencia del estatus social, económico y cultural de las familias en el rendimiento de los alumnos, se ha elaborado un índice estadístico que refleja ese estatus (ESEC) teniendo en cuenta el nivel de educación de los padres, la profesión y los recursos educativos domésticos.

En cuanto a las variables de esfuerzo, el estudio indaga sobre las actitudes de los alumnos hacia el estudio y la tasa de idoneidad. Finalmente, las variables del centro escolar se refieren a datos como la propiedad del colegio, el porcentaje de repetidores, número de profesores, agrupación de alumnos, información que ofrece el centro, presión de los padres, etc. Con estos y otros recursos, el informe PISA trata de encontrar explicación a las diferencias entre países, regiones y centros escolares, buscando, por tanto, los factores de éxito y fracaso. En el prólogo del informe español del PISA 2006, se expone el principal objetivo que anima a estos estudios impulsados por la OCDE: “Contribuir al mejor conocimiento de los aspectos fundamentales del funcionamiento del sistema educativo, analizar qué explican los resultados obtenidos y, sobre todo, facilitar la adopción de las políticas y las acciones educativas que permitan mejorar el sistema

educativo español” (MEC, 2007, p. 12); es decir, establecer una relación productiva entre el conocimiento y la actuación para la mejora de la educación. Christian Maroy y Catherine Mangez (2008) consideran que la relación entre el conocimiento educativo –a partir de la información que proporciona PISA y otras pruebas similares– sobre los problemas de la educación y el desarrollo de políticas educativas forma parte de un movimiento más amplio que pretende hacer ver las decisiones políticas como decisiones técnicas:

La promoción de encuestas a gran escala no puede ser disociada del contexto de la evolución de las relaciones entre conocimiento y política. Por otra parte, en relación con la incidencia que tienen estos estudios internacionales comparados a nivel de la globalización de las políticas educativas, es importante señalar que la difusión de estas encuestas participan de un movimiento más amplio que pretende una forma de racionalización de la acción pública en materia de educación. (Maroy y Mangez, 2008, p. 80).

Es decir, una de las ideas que domina el discurso de la política educativa actual es la de desarrollar medidas basadas en evidencias empíricas proporcionadas por las evaluaciones nacionales o internacionales. Se trataría así de salir de una especie de edad de piedra de las políticas educativas que, muy a menudo, eran resultado de teorías pedagógicas famosas reducidas al rango de modas, para promover una estrecha vinculación entre una investigación educativa de más calidad y decisiones más fundamentadas en los conocimientos empíricos, con resultados tangibles, y la constatación de probadas relaciones de causalidad. De esta forma la gobernación de los sistemas educativos se tiñe de un aura de objetividad aparentemente incontestable, pues, según esto, la base de las decisiones es de carácter científico y objetivo.

Para los citados autores, el discurso de los promotores de PISA va más allá. No se trata sólo de fundamentar la política educativa y las propuestas de mejora de los cen-

⁴ Al autor de estas líneas no le queda claro la virtualidad de esta clasificación de las variables explicativas, toda vez que existe una gran interrelación entre ellas y en buena medida los grupos b y c pueden considerarse dependientes de a.

tros, sino que, frente al conocimiento pedagógico calificado de *abstracto*, se trata también de estimular un tipo de investigación que conecte directamente con los objetivos de mejora que implícitamente han definido las mismas pruebas. La mejora de la educación debe valorarse en función del efecto que las políticas y actuaciones tienen sobre el aprendizaje de los alumnos, no sobre si “son más felices o más comunicativos” (Hextal y Mahony, 2001, p. 186).

La recepción en España de las pruebas de diagnóstico. El caso de Andalucía

Aunque ya en la Ley General de Educación de 1970 (BOE de 6 de agosto de 1970) se contemplaba la evaluación del sistema educativo, fue la Ley Orgánica de Calidad de la Educación (LOCE, BOE de 24 de diciembre de 2002) la que estableció por primera vez en España la realización de una evaluación o prueba de diagnóstico en primaria y secundaria, evaluación que versaría sobre competencias básicas del currículo (Art. 97); posteriormente, presentándose en el preámbulo como una novedad que no era tal, la Ley Orgánica de la Educación (LOE, BOE de 4 de mayo de 2006) hizo lo propio en sus artículos 22, 29 y 144, en los que se establecía que los centros realizarán una evaluación de diagnóstico⁵ de las competencias básicas alcanzadas por sus alumnos al finalizar el segundo ciclo de la educación primaria y el segundo curso de la educación secundaria obligatoria.

En este y otros asuntos, el caso de Andalucía es paradigmático de las políticas educativas neoliberales. Efectivamente, muy poco después de publicada la LOE, apenas días,

en una Orden de 28 de Junio 2006 (BOJA de 4 de agosto de 2006), la Consejería de Educación de la Junta de Andalucía regulaba la realización de la evaluación de diagnóstico en el ámbito territorial de su competencia, convirtiéndose en la primera Comunidad Autónoma que ponía en marcha el citado mecanismo⁶. De la misma forma que en la LOE, en la legislación andaluza se establece que, al finalizar el segundo ciclo de la educación primaria y el segundo curso de la educación secundaria obligatoria, los centros realizarán una evaluación de diagnóstico de las competencias básicas alcanzadas por su alumnado. Esta evaluación, se añade, tendrá un carácter orientador y formativo para los centros e informativo para las familias y para el conjunto de la comunidad educativa, descartándose implícitamente que los resultados puedan ser utilizados para establecer comparaciones o rankings entre los centros. Más concretamente, siguiendo el modelo PISA, anualmente se evaluarán alguna o varias de las llamadas competencias básicas. De hecho, en la primera evaluación realizada en Octubre de 2006, las pruebas versaron sobre la competencia en comunicación lingüística y de razonamiento matemático, al igual que en la de 2007, mientras que en 2008 se incorporó la competencia en el conocimiento e interacción con el mundo físico y natural. La realización de las pruebas tiene carácter censal, es decir, se hace a todos los alumnos y alumnas de centros públicos y privados concertados⁷, es preparada por equipos de especialistas y se materializa en unos cuadernillos donde los alumnos escriben las respuestas a las preguntas. La corrección corre a cargo del profesorado del mismo centro, si bien en una muestra (prueba de contraste) la realización y corrección fue realizada por personal ajeno, normalmente inspectores de educación.

⁵ Nótese que en esta literatura los términos *evaluación de diagnóstico* y *prueba de diagnóstico*, se utilizan indistintamente, aunque lo más habitual es que se diga evaluación cuando realmente se trata, sobre todo, de una prueba.

⁶ El proceso que se ha seguido en las demás Comunidades Autónomas ha sido muy similar. Aquí se da cuenta del caso de Andalucía al haber sido objeto de una investigación exhaustiva, realizada por este autor durante el curso 2009-2010 gracias a una licencia de estudios concedida por la Consejería de Educación de la Junta de Andalucía.

⁷ A partir de 2010 se realizará también al alumnado de los centros privados no concertados.

La puntuación obtenida por cada alumno se traslada a una escala 500 (Puntuación Transformada) que es el valor medio de todos los alumnos de Andalucía, teniendo en cuenta la Desviación Típica en los resultados obtenidos en las pruebas de cada una de las competencias. Por otra parte, siguiendo también el modelo PISA, se realiza un *questionario de contexto* que es respondido por el alumnado, tutores y familias. Con él se pretende recabar información sobre el contexto personal, familiar y escolar; así, se establecen una serie de variables suponiendo que están asociadas al nivel de competencia que alcanza el alumno; en las evaluaciones de 2006 y 2007 este cuestionario se realizó en una muestra de centros, mientras que en 2008 se generalizó a todos los de Andalucía. Con estos datos se construye un índice –Índice Socioeconómico y Cultural, ISC– para cada centro, que, junto a otras informaciones sobre sus características, puede ser utilizado para el análisis de los resultados⁸. Finalmente, añadir a esta descripción que a principios del curso 2009-2010 la Agencia Andaluza de Evaluación Educativa (AGAEVE) ha elaborado un nuevo indicador –*valor esperable*– que refiere al resultado que deberían obtener en conjunto los alumnos de un centro si se contextualiza la puntuación realmente obtenida en la media de su grupo socioeconómico y cultural.

La aplicación en España de las pruebas o evaluaciones de diagnóstico y, más concretamente, en Andalucía, es en realidad una mera traslación del discurso que más arriba se ha esbozado para PISA. Tanto en los textos legales que regulan el mecanismo como en los informes elaborados cada año por la administración educativa se repite el mismo esquema:

[A partir de los resultados] Los centros, a la vista de sus circunstancias y características, tras un proceso de análisis y reflexión riguroso, establecerán las propuestas de mejora del rendimiento del alumnado que consideren necesarias. Se trata de una información que, utilizada en su justa medida, pueda ser útil al centro, al profesorado, al alumnado y a las familias para coordinar esfuer-

zos en la mejora del rendimiento escolar. No hay que olvidar tampoco el carácter eminentemente orientador que poseen dichas pruebas, por cuanto permiten conocer la evolución del rendimiento del alumnado a lo largo de los próximos años, valorando el efecto que sobre el mismo puedan tener las propuestas de mejora introducidas. (BOJA de 4 de agosto de 2006).

Es decir, la información que se obtiene de las pruebas de diagnóstico proporciona el combustible para el funcionamiento de la dinámica de la autoevaluación, trasladando al centro la responsabilidad de la puesta en marcha de medidas que mejoren los resultados, medidas cuya efectividad volverá a ser evaluada en las pruebas del curso siguiente. Evaluar para conocer y mejorar es, precisamente, el lema que reza en el frontispicio de la AGAEVE. Además, esa misma información –supuestamente rigurosa y objetiva– es la base sobre la que la administración determina su política educativa:

Esta evaluación, en la que participan unos 200.000 alumnos y alumnas y más de 4.000 centros de Educación Primaria y ESO, se ha concebido con una doble finalidad: por un lado, ayudar a los centros educativos a mejorar sus prácticas docentes y el rendimiento del alumnado y, por otro, facilitar a la Administración educativa información relevante para orientar la política educativa (Benítez, 2010, p. 84).

Los supuestos teóricos del discurso sobre las evaluaciones y pruebas de diagnóstico

Más allá de su apariencia de rigor y científicidad, el sentido, la relevancia y centralidad que han adquirido las evaluaciones y pruebas de diagnóstico tienen mucho que ver con el despliegue de políticas neoliberales. En el campo de la educación estas políticas se han nutrido con diversos ingredientes y líneas de pensamiento. Es el caso, en primer lugar, del llamado movimiento de la *Eficacia Escolar*. De alguna forma, el Informe

⁸ En el caso de la muestra utilizada en nuestro estudio (los centros de Educación Secundaria públicos y privados concertados de la provincia de Sevilla), los valores del ISC oscilan entre 1,44 y 1,42. El conjunto se distribuye en cuatro grupos: Bajo, Medio-bajo, Medio-alto y Alto.

Coleman⁹ está en el origen de este movimiento de carácter más bien teórico y centrado en la investigación. Mientras que en ese informe se subrayaba la importancia del contexto socioeconómico y cultural como factor determinante del rendimiento de los alumnos, algunos investigadores llamaron la atención sobre el hecho de que parte de los diferentes resultados que obtenían los centros de contexto similar podría deberse a la escuela y a sus profesores. A partir de ahí se suscitaba una cuestión fundamental: cuáles son los factores que están presentes en las escuelas de éxito o eficaces. Naturalmente ello requería una definición previa acerca del concepto de eficacia, definición que de una u otra forma acaba polarizándose en los resultados que obtienen los alumnos y en otros indicadores que resulten fácilmente mensurables. Así, la escuela eficaz es aquella que en un contexto dado obtiene mejores resultados. Cuando esos resultados están por encima de la media de su contexto, se habla de “valor añadido”, es decir, de la capacidad que tienen la escuela y los profesores de mejorar con su trabajo los resultados de los alumnos de contextos similares.

Por otra parte, amparándose en el supuesto de que existe una racionalidad científico-técnica capaz de conseguir escuelas eficaces mediante la aplicación de un instrumental apropiado, la perspectiva de la eficiencia pretende emular lo que ocurre en el campo de las ciencias positivas y de la técnica. Efectivamente, la convergencia entre el movimiento de la eficacia escolar y el que se ha dado en llamar de mejora de la escuela, permitía establecer relaciones entre los estudios sobre los factores de las escuelas eficaces y las estrategias sistemáticas de mejora. Mientras que

el movimiento de escuelas eficaces se había centrado en la identificación de los factores de éxito y fracaso y el de la mejora lo hacía en el cambio de la escuela, el refundido, bautizado como movimiento de *Mejora de la Eficacia Escolar*, subrayaba el vínculo entre la investigación, las iniciativas de mejora y la política educativa. Uno de los núcleos fundamentales de este discurso serán las evaluaciones o pruebas de diagnóstico. Como ya se ha dicho, mediante este recurso se podrá disponer de indicadores de éxito y fracaso, así como conocer los factores que producen los resultados y, por ello, saber las medidas que deben adoptarse para mejorar los resultados.

¿Qué conocimiento proporcionan realmente las evaluaciones o pruebas de diagnóstico?

Las evaluaciones de diagnóstico tratan de obtener resultados y explicaciones a las diferencias entre centros, países o regiones, buscando, por tanto, los factores que hacen posible la eficacia escolar, es decir, los buenos resultados de los alumnos. A este respecto distintos estudios, incluso los elaborados por las instituciones y administraciones patrocinadoras de los programas de evaluación, llegan a conclusiones bastante similares: el factor socioeconómico y cultural es el factor más decisivo en el rendimiento de los alumnos¹⁰. Concretamente, en un trabajo sobre los datos de PISA 2006, Carabaña (2009) afirma que ni el sistema educativo ni la gestión pública o privada de los centros escolares ni otras de las variables explicativas consideradas inciden de manera significativa en los resultados:

⁹ Este informe, publicado en 1966 con el título de *Equality of Educational Opportunity Study*, implicó a 600.000 alumnos y 60.000 docentes de 4.000 centros educativos de Primaria y Secundaria de Estados Unidos. Fue resultado del importante estudio realizado por James S. Coleman y su equipo de investigadores, por encargo del Congreso de los Estados Unidos, tras la aprobación, en 1964, de la Ley de Derechos Civiles, y respondía al mandato de conocer si todos los alumnos tenían las mismas oportunidades educativas, independientemente de su grupo étnico-racial, color, religión u origen socio-cultural.

¹⁰ Con respecto a los datos de PISA 2006, así se afirma, entre otros, en el Informe español PISA 2006 (MEC, 2007), en el estudio de López Martín y otros (2009) y en el de Julio Carabaña (2009). En el mismo sentido se pronuncia el estudio de García Pérez, Pinto y Robles (2009). Por lo demás, los informes de la administración andaluza sobre las pruebas de diagnóstico (Evaluaciones de diagnóstico, 2006-2007; 2007-2008 [avance] y 2008-2009) llegan también a la misma conclusión.

Los estudios PISA permiten afirmar que las diferencias entre los países no se deben a las características de sus escuelas, ni a nivel de sistema ni a nivel de centro. El único factor que tiene una influencia importante y consistente sobre las diferencias de aprendizaje entre países es la composición social de sus poblaciones. Si esta composición fuera la misma en todas partes, sus puntuaciones medias en las pruebas PISA se igualarían mucho, quedándose las diferencias en casi la mitad. Por desgracia, este único factor influyente no es susceptible de modificación por medidas políticas, al menos a corto plazo. [...] Por este camino de los modelos estadísticos se llega a la conclusión de que ni los recursos materiales, ni la ordenación general de los sistemas, ni los modelos de gestión, ni las prácticas pedagógicas pueden explicar una mínima parte de las diferencias entre países. Ello por dos razones. Una, que no tienen una influencia consistente a nivel de centro; dos, que, incluso aunque la tuvieran, esta influencia no se trasladaría a nivel de país. El único factor con poder explicativo importante es la composición sociocultural de las poblaciones, un rasgo de los individuos que le viene de fuera a la escuela. (Carabaña, 2009, pp. 92-93).

Efectivamente, detrayendo el ESEC, es decir, desactivando la incidencia del contexto en los resultados mediante una operación aritmética, resulta que “en los niveles socioeconómicos más modestos, los alumnos españoles obtienen mejores resultados, o, lo que es lo mismo, en los centros que escolarizan a estos alumnos el sistema educativo ofrece mejores resultados que en los alumnos equivalentes de la OCDE” (ibíd.).

El caso es que los estudios sobre los resultados de las pruebas de diagnóstico realizadas en Andalucía en 2006, 2007 y 2008 llegan también a conclusiones similares. Los informes elaborados por la Consejería de Educación sobre las pruebas realizadas en octubre de 2006 y, especialmente, el de 2008-2009 (Evaluaciones de diagnóstico, 2006-2007; 2007-2008 (avance) y 2008-2009) analizan la relación que existe entre distintas variables y los resultados que obtienen los alumnos. A este respecto, se afirma que: “Los niveles competenciales [...] se incrementan con el ni-

vel socioeconómico y cultural del alumnado. Las pruebas de contraste realizadas revelan que estas diferencias resultan ser estadísticamente significativas” (Informe 2006-2007, p. 69). Igualmente se constata una relación directa de los resultados con otros indicadores de contexto que son claramente dependientes del anterior: estudios alcanzados por los padres y madres, expectativas de los padres y madres respecto al nivel educativo que logrará su hijo o hija, tiempo que dedican sus padres y madres a la lectura y número de libros presentes en los hogares¹¹. Así mismo, se observa también esa misma relación entre los resultados y otra serie de variables que el informe denomina *rasgos, comportamientos y actitudes personales*, como son el género –“las alumnas superan a los alumnos en la Competencia en Comunicación Lingüística (tanto en Primaria como en Secundaria), mientras que los resultados en Competencia Matemática favorecen a los alumnos en ambas etapas” (ob. cit., p. 29)-, la tasa de idoneidad –“la mera repetición de curso no afecta al aumento del rendimiento de los alumnos y alumnas que han cursado más de un año en el mismo nivel educativo” (ob. cit., p. 33)-, la dedicación a la lectura –“Se aprecia una tendencia a que cuanto más elevada sea la puntuación en el factor de lectura mayor sea la media lograda en ambas competencias, tendencia que resulta algo más acusada en el caso de la competencia en comunicación lingüística” (ob. cit., p. 39)- y el tiempo dedicado al estudio (ob. cit., p. 43), factores todos ellos que, exceptuando el caso del género, son igualmente dependientes en su mayor parte del contexto socioeconómico y cultural de los alumnos¹².

Similares conclusiones se expresan en los informes sobre los resultados de las pruebas de 2007 y 2008, sin que tres años después del primer estudio se adviertan novedades significativas.

Por nuestra parte, la investigación que hemos realizado sobre los datos de las tres pruebas realizadas en Andalucía nos pone

¹¹ A este respecto puede verse también Gil Flores, 2009.

¹² Por lo demás, el factor género es una variable cuyo comportamiento no puede variarse fácilmente con actuaciones desde los centros escolares. En todo caso, las diferencias en el comportamiento de hembras y varones respecto al estudio también tiene mucho que ver con factores de orden histórico y sociocultural.

de manifiesto igualmente que el factor socioeconómico y cultural es el más influyente en la determinación de los resultados que obtienen los alumnos, mientras que otras variables analizadas –como el número de programas educativos de los centros, la existencia o no de Planes de Autoevaluación y Mejora, de Planes de Lectura y Biblioteca o el tipo de enseñanzas que imparten– no tienen una incidencia especialmente significativa en el rendimiento de los alumnos.

En fin, en el estudio realizado por la Consejería de Educación de Asturias sobre pruebas de diagnóstico realizadas en 2006 (Pérez Cullera, 2008) se llega a conclusiones parecidas a las expuestas anteriormente: “Como era esperable el índice SEC [relativo al contexto socioeconómico y cultural] del alumnado está positiva y significativamente relacionado con el resultado en la prueba: a mayor nivel social, mejor puntuación” (ob. cit., p. 56). Por lo demás, precisamente en este estudio es posible advertir con mayor claridad el sesgo intencional con el que se elaboran. Efectivamente, en la línea de sobredimensionar el llamado *efecto escuela*, los autores se ocupan de subrayar que: “Es cierto que algunas variables de entrada, en especial el nivel sociocultural del centro, inciden en los resultados del alumnado. Sin embargo, como se verá más adelante, el efecto de variables como las expectativas, la motivación y el autoconcepto académico tienen una influencia mucho más decisiva en los resultados” (ob. cit., p. 54), ignorando deliberadamente que esas variables son a su vez dependientes en gran medida del nivel sociocultural. A falta de un mínimo rigor en estas y otras afirmaciones, el argumento se justifica con razones previas al estudio, anunciándonos algunas de los objetivos ocultos de las evaluaciones de diagnóstico: “Ello permite concluir que los centros, por medio de la intervención sobre factores personales y hábitos de trabajo, disponen de un margen de maniobra impor-

tante para la mejora y compensación de las desigualdades ante la educación” (ob. cit., p. 54), como si los factores personales y hábitos de trabajo fueran ajenos a la identidad histórica, social y cultural de los alumnos.

Así pues, tal y como hemos visto en los casos analizados, la información que proporcionan las evaluaciones de diagnóstico nos llevan a la conclusión de que el factor más relevante en la explicación de los resultados de las pruebas es el del contexto socioeconómico y cultural, mientras que otros factores tienen una incidencia mucho menor y muy desigual. Si el objetivo de las evaluaciones y pruebas de diagnóstico es el de facilitar información y producir conocimiento sobre la mejora de la educación (o de los resultados), hay que decir que realmente nos dicen muy poco (poco más de lo que ya sabemos) sobre las diferencias de rendimiento y nada sobre lo que habría que hacer para mejorar los resultados (Carabaña, 2009).

Acerca del “valor añadido” y el oportunismo del llamado “efecto escuela”

Si bien no está clara la incidencia de su esfuerzo en los resultados que obtienen los alumnos, de lo anterior no debe deducirse que para la formación de niños y jóvenes sea irrelevante lo que hacen los centros y los profesores. Desde luego la influencia del contexto sobre el rendimiento no es determinante de manera absoluta; de hecho en todos los estudios se destaca, en términos estadísticos, la existencia de un grupo de alumnos (en torno al 10%) cuyos logros están por encima de lo que se corresponde con sus indicadores socioeconómicos y culturales¹³. Puede pensarse que, en estos casos, los centros y los profesores *añaden valor* en el proceso de enseñanza y aprendizaje, pues consiguen que sus alumnos obtengan mejores resultados de los que

¹³ En el estudio que hemos hecho sobre las pruebas realizadas en Andalucía, los centros de ISC bajo y medio-bajo cuyos resultados se sitúan por encima de la media de puntuación constituyen el 10% del total. Un porcentaje similar es el que se constata en el informe asturiano citado más arriba (Pérez Cullera, 2008, p. 49). En el caso de los centros con ISC medio-alto y alto, este porcentaje se sitúa en torno al 60%, lo que da pie a pensar que la incidencia del contexto sociocultural es más determinante en los niveles más bajos.

obtendrían en otro centro y con otros profesores situados en la media de la escala. Desde la perspectiva de la *eficacia para la mejora escolar*, este tipo de centros se convierte en su principal objeto de estudio, pues, frente a lo que se decía en el informe Coleman, demostrarían que *la escuela sí importa*. Así mismo, tal y como se ha señalado más arriba en el caso asturiano, las políticas educativas inspiradas en este planteamiento también se aferran a esa idea: independientemente del contexto, los centros y los profesores pueden hacer mucho para mejorar los resultados de los alumnos; por lo tanto, la política educativa debe centrarse precisamente en esto. He aquí cómo, de forma casi inadvertida, el problema del rendimiento de los alumnos ya no es un problema del contexto sino de los centros y profesores.

Los trabajos críticos con estos planteamientos han subrayado la debilidad conceptual de sus supuestos, las limitaciones de las pruebas como fuente de información y el oportunismo de las políticas que los utilizan¹⁴. Si, como se ha dicho, en términos generales añaden poco a lo que ya sabemos, las evaluaciones de diagnóstico no nos aportan información relevante acerca de las características de esos centros *eficaces* ni de los factores que explican sus resultados excepcionales; tampoco nos dicen mucho acerca de las peculiaridades de los centros *ineficaces*. Además, aunque pudiéramos conocer con detalle el perfil y las circunstancias de unos y otros, no es fácil ir más allá de meras correlaciones de dudosa solvencia explicativa. Por otra parte, dada la importancia (por su fuerte incidencia) de los contextos locales y de las culturas específicas de los sistemas y de los centros educativos, no está claro que lo que en un centro contribuye al éxito escolar, opere de la misma forma en otro. Igualmente dudoso es que, si admitiéramos que existen unos factores de éxito, puedan reproducirse en las escuelas *ineficaces*, entre otras razones porque generalmente éstas no funcionan de forma contraria a las otras. También es más

que discutible que, permaneciendo constante la variable fundamental –la incidencia del contexto socioeconómico y cultural–, puedan conseguirse cambios significativos actuando sobre variables poco relevantes. En fin, sustentar la política educativa y las estrategias de mejora (de los resultados) sobre lo que ocurre en el 10% de los centros escolares, ignorando lo que sucede en el 90%, quizás no se deba a razones científico-técnicas, sino a intereses de otro tipo.

Por otra parte, las críticas a los supuestos en los que se apoya la práctica de las evaluaciones se han dirigido también al propio instrumento de recogida de la información. En lo que respecta a los contenidos, validez científica de las pruebas y uso estadístico de los resultados, se han planteado interrogantes que cuestionan su rigor. Por una parte, el enfoque de competencias no está exento de discusión, tanto en lo que respecta al sentido que le confiere a la enseñanza como al modo en que se evalúan. Por ejemplo, Jones y Moore (2008) consideran que se trata de una interesada desvirtuación de conceptos de la pedagogía progresista, puestos al servicio de la política cultural de la Nueva Derecha y de los enfoques de la educación como capital humano. Crahay (2006) cuestiona la relación entre conocimiento y competencias y, sobre todo, el hecho de que las pruebas estandarizadas puedan efectivamente medir el grado de adquisición (vid. también Mons, 2009). Gimeno (2008a y 2008b) plantea cómo la naturaleza política del concepto aplicado a la educación –en relación con la estrategia de Lisboa y la apuesta por una política educativa de la Unión Europea– lo convierte en un instrumento artificioso y de difícil viabilidad para articular los contenidos curriculares. En fin, Goldstein (2008) cuestiona también la idoneidad de las pruebas homogéneas, considerando que es necesario tener en cuenta las especificidades culturales en la confección de las preguntas; así mismo llama la atención sobre las operaciones estadísticas que se ha-

¹⁴ Por razones de espacio no vamos profundizar en esos trabajos. Son numerosos los estudios en los que se pueden encontrar los argumentos críticos en relación con la perspectiva de la eficacia y con las políticas conservadoras y social-liberales que la han puesto en marcha. El lector o lectora puede encontrar dos buenas síntesis en Wrigley, 2007 y Slee y otros, 2001.

cen sobre los resultados, entendiendo que, frente a lo que es habitual, los modelos estadísticos deben reflejar la complejidad de la realidad, de manera que la multidimensionalidad de los fenómenos estudiados y las diferencias culturales deben ser conservados en lugar de ser eliminados a favor de una escala común. A ello debemos añadir que las comparaciones de resultados obtenidos con pruebas distintas cada año y realizadas a alumnos distintos no es un procedimiento fiable.

En definitiva, las evaluaciones y pruebas de diagnóstico no constituyen un recurso infalible ni indiscutible (ni, por supuesto, objetivo) para diagnosticar los problemas de la educación. No nos dicen nada que no sepamos acerca de los factores más influyentes en el rendimiento escolar, no nos aclaran cuáles son las circunstancias que explican el caso de los centros escolares que tienen unos rendimientos superiores a la media de su contexto y, desde luego, no nos explican qué habría que hacer para que los centros *ineficaces* alcanzaran mejores resultados. Además, como se recoge en distintos informes, generalmente los resultados coinciden con las calificaciones que obtienen los alumnos en las evaluaciones curriculares. Es decir, las evaluaciones y pruebas de diagnóstico nos proporcionan muy poca información sobre la mejora de la educación, de manera que su existencia no se justifica por el fin que dicen perseguir, sino por razones de rentabilidad y oportunidad política, ya que forman parte de un dispositivo más amplio inscrito en el marco de las políticas conservadoras y social-liberales.

La política educativa no es resultado de las evaluaciones de diagnóstico.

Las evaluaciones y pruebas de diagnóstico son resultado de la política educativa

Hacia la globalización del currículum

Tröhler (2009) plantea la tesis de que estamos inmersos en un proceso de armonización del globo educativo, en una política educativa mundial, de la que PISA represen-

ta su más distinguido instrumento. La potencia de estas pruebas reside en la capacidad que tienen los poderes supranacionales (la OCDE y la UE) para extender, a través de ellas, el planteamiento de las competencias como referencia central en la formulación de la enseñanza en todo el mundo.

En el caso de las evaluaciones domésticas, el objetivo de diseñar desde fuera los contenidos y la práctica de la enseñanza, queda también patente en algunas declaraciones. En un artículo citado anteriormente (Benítez, 2010), se nos muestra con claridad este propósito. Al evaluar competencias y no conocimientos en sentido estricto, la realización de las pruebas tiene como objetivo producir cambios en los contenidos del currículum y de la práctica docente sin necesidad de cambiar el currículum oficial ni el discurso sobre los métodos de enseñanza, tal y como se planteaba con las reformas pedagógicas: “Esta concepción conlleva un cambio en las prácticas docentes, ya que se trata no sólo de enseñar conocimientos, sino de enseñar para la vida. Podemos afirmar que, en este contexto, los contenidos curriculares tal y como se han venido concibiendo hasta ahora por algunos sectores, es decir, como objetivos en sí mismos, tienen fecha de caducidad” (Benítez, 2010, p. 83).

Determinar la agenda de la investigación educativa

En un artículo titulado significativamente *PISA et les enquêtes internationales. Enjeux scientifiques, enjeux politiques*, Jean-Yves Rochex (2008) considera que las intenciones y razones políticas de PISA son claras, pero que quizás se disimula otra menos explícita como es la de contribuir a una transformación en profundidad de la investigación educativa a escala internacional y en los diferentes países. Por una parte entiende que la profusión de este tipo de pruebas, a nivel internacional y en el interior de cada país, obedece a la puesta en marcha de políticas educativas basadas en los modelos de cuasi-mercado y en el Estado-evaluador. Es decir, no es la información que proporciona PISA la que, como se dice, genera políticas

y actuaciones de los gobiernos, sino que las políticas educativas dominantes han producido PISA dentro de una lógica de evaluación, rendición de cuentas y control de la profesión docente. En este sentido, los instrumentos de evaluación se convierten en técnicas de cambio y los investigadores en expertos que proporcionan las evidencias; pero, puesto que la investigación es incapaz de ofrecer resultados consistentes y fiables, el discurso político tiende ahora a modelar la investigación educativa, determinando objetos de estudio –los factores de éxito y los resultados– y métodos de investigación –los cuantitativos frente a los cualitativos–.

Legitimar la política educativa

Ya hemos dicho que Maroy y Mangez (2008) cuestionan la afirmación tantas veces reiterada en las declaraciones oficiales de que las evaluaciones contribuyan a la mejora de la educación y de la política educativa, preguntándose si los datos que proporcionan las pruebas sirven verdaderamente para racionalizar la acción pública o lo que realmente se produce es una politización del conocimiento, es decir, un uso político de esos datos. En un estudio sobre las reacciones de distintas instituciones, vertidas en la prensa a propósito de las pruebas, los autores constatan cómo, efectivamente, en el caso de Bélgica, los mismos resultados PISA se utilizan para justificar, argumentar y apoyar tres discursos distintos: el de la ministra de educación, el del movimiento socialista APED y el de la Red de Escuelas Católicas.

Siguiendo esta pista, hemos hecho una breve indagación acerca de los titulares de prensa que generaron en España los resultados PISA 2006 y otras evaluaciones de diagnóstico. A modo de ejemplo vale decir que el 18 de noviembre de 2006, el diario *El País* titulaba una información sobre PISA 2006 con el siguiente texto: “El contexto social influye más en las notas que el tipo de colegio” (*El País*, 18 de noviembre de 2006), mientras que, en ese mismo diario, a propósito de los resultados de la evaluación general de diagnóstico realizada en España en 2010, se afirmaba: “La clave del éxito: el colegio. El

último estudio sobre los resultados escolares en primaria revela la importancia de la autonomía del centro. La flexibilidad bajo control se apunta como solución” (*El País*, 17 de Junio de 2010). El contenido de la información periodística aludida discurre entre alabanzas de las autoridades a la autonomía de los centros y otras modas por el estilo. Las pruebas de diagnóstico al servicio de una política educativa: “Hemos visto que lo importante es lo que ocurra dentro del centro”, asegura el director del Instituto de Evaluación, encargado del estudio, Enrique Roca. Es decir, en sólo algo más de tres años, el contexto social ha dejado de ser el factor relevante en los resultados, mientras que lo que antes influía menos, el colegio, ahora resulta que, gracias a los resultados, es lo que importa más.

Las evaluaciones de diagnóstico en los dispositivos de control de los centros y de los docentes: incentivos y rendición de cuentas

A la vista de lo anterior, parece claro que las evaluaciones y pruebas se ponen al servicio de una nueva forma de gestión y control de los centros escolares y de la batalla por los resultados: la rendición de cuentas, los incentivos y la autoevaluación constituyen, junto al gerencialismo, los elementos fundamentales de esa estructura.

La deriva del papel del estado en la política educativa neoconservadora, desde una posición en la que dirigía los procesos de reforma hasta la actual en la que se limita a evaluar, sitúa los resultados que obtienen los alumnos en un plano distinto al del diagnóstico. En países como Inglaterra y Estados Unidos, desde hace ya algunos años, las pruebas estandarizadas juegan un papel importante en la financiación de los centros escolares, en las condiciones de trabajo de los docentes y en la competencia entre escuelas. Los resultados que obtienen los alumnos constituyen una pieza fundamental en la política de incentivos, en el establecimiento de *rankings* y en la existencia misma de los centros escolares. Esta aparente transmutación del sentido de las pruebas estandarizadas ha hecho que se conviertan en un “instrumento idóneo para poner en tela de

juicio el trabajo del docente. El examen, con su aplicación masiva, ha pasado de ser un instrumento del docente, para convertirse en un instrumento contra el docente” (Padilla, 2009, p. 45).

En España, sin llegar, de momento, a generalizarse la relación entre financiación de los centros y resultados¹⁵, se ha planteado algo parecido, concretamente en la comunidad autónoma andaluza, con la aplicación del Programa de Calidad y mejora de los rendimientos escolares¹⁶. El Programa es, donde los haya, un ejemplo paradigmático de la aplicación de la estrategia de incentivos económicos para la mejora de la educación, entendida como consecución de resultados centrados en el rendimiento de los alumnos. Importada claramente del fordismo industrial, la fórmula y la primera convocatoria se materializaron en una Orden de 20 de febrero de 2008, publicada en el BOJA el 29 de ese mismo mes y año. Resumidamente consistía en pagar a los docentes que se comprometieran unos extras económicos en función de la consecución de una serie de objetivos entre los que se encontraba –y con un importante peso sobre el total– la mejora de los resultados de los alumnos en las pruebas de diagnóstico.

Las evaluaciones de diagnóstico en los dispositivos de control de los centros y los docentes: la autoevaluación

Como ya se ha dicho, las pruebas de diagnóstico constituyen también una pieza del mecanismo de la autoevaluación que las

políticas educativas conservadoras publicitan de manera persistente como estrategia de mejora. En diversos círculos se destaca la importancia de la “cultura de la evaluación” para mejorar, y hoy es una plaga que se ha extendidos por todos los ámbitos de la gestión pública, incluido el campo de la educación. El mecanismo de la evaluación y autoevaluación requiere el establecimiento de indicadores y de recursos para conocerlos. La información que proporcionan esos indicadores, especialmente los resultados que obtienen los alumnos, se constituye en el punto de partida del proceso de mejora: se contrasta la diferencia entre los objetivos perseguidos y los resultados obtenidos, se analiza qué se ha hecho mal y qué se ha hecho bien (puntos fuertes y puntos débiles) y se traza un plan para corregir los errores y mejorar los resultados.

Esta estrategia, basada meramente en el sentido común, no sólo no ha producido resultados empíricamente demostrables¹⁷, sino que adolece de cierta ingenuidad¹⁸, aunque al mismo tiempo entraña alguna dosis de perversidad. En realidad, se trata de una fórmula importada del mundo de la empresa, de un modelo de gestión postfordista (al estilo de los modelos conocidos por las siglas EFQM) que trata de lograr el compromiso de los trabajadores con las metas de la organización, produciendo entornos autónomos y flexibles (Bolívar, 2001, p. 7) que les impliquen –y, añadimos, les responsabilicen– en la consecución de unos objetivos.

La autoevaluación se interpreta como un dispositivo de control e interiorización de la

¹⁵ En cierta medida los Planes de Autoevaluación y Mejora han constituido una punta de lanza de esta estrategia. Por otra parte, recientes declaraciones de autoridades educativas apuntan a la fórmula de *contrato-programa* como sistema de financiación ligado a resultados.

¹⁶ Desde hace unos meses el citado programa –que tuvo escasa acogida especialmente en los Institutos de Secundaria– está derogado. Además, a la hora de redactar estas líneas el Tribunal Superior de Justicia de Andalucía lo ha declarado ilegal.

¹⁷ La investigación que hemos desarrollado sobre los Planes de Autoevaluación y Mejora en Andalucía nos permite afirmar que la dinámica de la autoevaluación no tiene incidencia significativa sobre los resultados que obtienen los alumnos en las pruebas de diagnóstico ni en las calificaciones ordinarias.

¹⁸ Así, por ejemplo, pedir a los centros que, a partir de unos datos, formulen propuestas de mejora, es como pedir a la meteorología que a partir de una serie anual de precipitaciones nos diga qué debemos hacer para que llueva más o menos.

culpabilidad ante los previsible fracasos, una tecnología del poder que permitiría a la vez gobernar a los centros escolares y a los profesores y hacerlos responsables de los resultados de los alumnos. En el marco del reiterativo discurso sobre la autonomía, la autoevaluación es un recurso de extraordinaria potencia para resolver los conflictos que se generan en el desarrollo de la política educativa. En este sentido Malen (1994, p. 251, citado en Bolívar, 1994, p. 919) ha afirmado que, a la vista de la débil relación que existe entre las iniciativas gubernamentales y la mejora de la calidad educativa y del aprendizaje de los alumnos, su persistencia se explica más bien por la utilidad política que aportan. En la estrategia de retener el control del sistema, propagando a la vez el discurso de la autonomía de los centros escolares –retórica descentralizadora y comportamiento centralizador–, el mecanismo de la autoevaluación otorga legitimidad a la intervención externa, pues da voz a los docentes –“aparentando que el que un centro funcione depende sólo del propio centro” (Bolívar, 1994, p. 922)–, aunque realmente se les diga lo que deben hacer y se les haga responsables de los resultados. Paradójicamente, el énfasis en la autoevaluación es reflejo de la nueva relación entre política y educación.

Los efectos reales de las evaluaciones y pruebas de diagnóstico

¿Realmente mejoran los resultados?

En una nota de prensa fechada el 18 de octubre de 2006, en la que se reseñaba la intervención de la Consejera de Educación en el Parlamento andaluz, decía la citada responsable que “las pruebas proporcionarán información ‘objetiva y rigurosa’ tanto para la mejora del rendimiento del alumnado como de la práctica educativa” (Nota de prensa, 2006). Cuatro años después de aquellas declaraciones, y tras cuatro pruebas de diagnóstico realizadas, no hay nin-

guna evidencia empírica, antes al contrario, de que, efectivamente, se haya producido una mejora del rendimiento del alumnado ni de la práctica educativa. De hecho, los resultados de las pruebas de diagnóstico de 2006, 2007 y 2008, en Andalucía, reflejan una tendencia a la baja. Por otra parte, aunque, como señala Mons (2009), los estudios sobre las políticas de rendición de cuentas y de evaluaciones externas son poco fiables, ya que adolecen de tales fallos metodológicos –y triquiñuelas (Mons, 2009, p. 112)– que en buena parte los invalidan, lo que sí queda claro es que el sistema de evaluación basado en el rendimiento de los alumnos en pruebas externas y en el uso de incentivos –premios y sanciones– a profesores y centros escolares no ofrece unos resultados suficientemente consistentes (Mons, 2009, p. 117); incluso el último informe PISA revela que no existe una relación estadísticamente significativa entre el rendimiento de los alumnos y la existencia de las pruebas de diagnóstico (OCDE, 2008, p. 253).

En el mismo sentido cabe referirse a otro de los objetivos de los dispositivos que estamos analizando, el de la reducción de las desigualdades escolares. Sobre este último punto, Wrigley (2007, p. 47) es concluyente: “El intento de hacer más eficaces las escuelas de las áreas más pobres por medio de una gestión hábil del cambio interno [...] ha llevado a un aumento de las diferencias de los resultados obtenidos entre los distintos grupos sociales”. Es decir, las diferencias de resultados académicos han aumentado en Inglaterra y Gales. Un dato que avalan otros estudios: “Cuanto mayores son los incentivos, el sistema de rendición de cuentas reduciría más la calidad de la educación en escuelas en las que las ganancias de rendimiento son más necesarias” (Ladd y Walsh, 2002, p. 16, citado en Ferrao, 2009, p. 140). En fin, los resultados de las pruebas de diagnóstico en Andalucía confirman también esta tendencia: el porcentaje de centros de nivel ISC bajo que supera la media de las puntuaciones disminuyó entre 2006 y 2008 (del 10% al 8%), mientras que aumentó el de los centros que se sitúan en el nivel social más alto (del 56% al 62%).

Los riesgos y “efectos colaterales” de las evaluaciones y pruebas de diagnóstico

Así pues, no puede decirse que los mecanismos de evaluación externa y rendición de cuentas hayan producido una mejora evidente y generalizada de los resultados que obtienen los alumnos, sino que, en no pocos casos, incluso ha generado una brecha mayor en los estudiantes de grupos sociales distintos, o, al menos no ha sabido reducirla. Pero, si en este campo son discutibles los efectos positivos, o, al menos, no quedan claros, en otros aspectos del funcionamiento del sistema educativo, de la dinámica de los centros escolares, de la práctica de la enseñanza y de la actuación de los profesores, la incidencia de los dispositivos de evaluación externa sí parece notarse. Sahron y Berliner (2007) hablan de los “efectos colaterales” de las pruebas estandarizadas para referirse a una serie de consecuencias negativas sobre la educación

En primer lugar, la realización de este tipo de pruebas induce a alumnos y profesores a centrar su atención en prepararlas, restringiendo el contenido de lo que se enseña y se aprende y ocupando tiempo en detrimento de otras tareas y conocimientos (Grant, 2000; Mons, 2009; Barrenechea, 2010). Los contenidos de la enseñanza se van centrando en los aspectos que van a ser objeto de examen, abandonándose otros conocimientos; mientras que la práctica de la enseñanza tiende a polarizarse en torno a la preparación del examen, empobreciéndose de esta forma el sentido de la educación.

Por otra parte, se ha subrayado la incidencia negativa en el comportamiento de los profesores y centros sobre los alumnos con más dificultades. Así, se tiende a prestar más atención a los alumnos que tienen posibilidades de obtener resultados satisfactorios, abandonando a su suerte a los que seguramente tendrán malos resultados. En muchos casos los directores sugieren a los “malos alumnos” que abandonen la escuela, al menos el día de la prueba, para no bajar los resultados (Wrigley, 2007; Mons, 2009). Además, cuando la remuneración y el estatus profesional están ligados a los resultados, los profesores tienden a abandonar

los centros con alumnos más difíciles para trabajar en otros con más posibilidades de éxito; es por esto por lo que Ladd y Walsh, citados más arriba, han constatado que la política de incentivos perjudica a los centros que tienen más necesidades de mejora, de manera que, como ya señalaba Wrigley, uno de sus efectos perversos es precisamente que aumenta las diferencias entre alumnos de grupos sociales diferentes.

Barrenechea (2010) advierte de la generalización de prácticas ilícitas con el fin de mejorar los resultados. El hecho de que tengan efectos en el prestigio y financiación de los centros, en los incentivos que perciben los profesores, así como la influencia de su repercusión mediática en la imagen de la administración educativa hacen que, en algunos casos, se “ayude a los alumnos” a responder correctamente, se baje el nivel de las pruebas o se maquillen estadísticamente los datos, lo que, desde luego repercute en el papel formativo de la escuela, y en realidad no mejoran el nivel educativo. Ciertamente se tiende a corregir estas “desviaciones”, pero, como afirma Barrenechea, mientras más importancia se le da a los resultados, más se agudiza el ingenio para buscar fórmulas que reduzcan sus efectos.

A este respecto, el caso de las pruebas estandarizadas que se realizan en los Estados Unidos de América es particularmente paradigmático. Como se sabe, la ley de 2001, conocida por sus siglas NCLB (*No Children Left Behind*), establecía un sistema de financiación extraordinaria del gobierno federal a los estados, en función de los resultados que obtuvieran los alumnos en las pruebas estandarizadas. En un reportaje periodístico publicado el 22 de Diciembre de 2000 en el diario *New York Times*, Jacques Steinberg (2000) informaba de cómo los malos resultados que obtenían los alumnos en las citadas pruebas estaban obligando a muchos estados a revisar a la baja el tipo de examen que realizan los estudiantes. De hecho, más recientemente, el presidente de aquel país, Barack Obama, a la vista de estas prácticas relativamente fraudulentas, ha planteado la necesidad de un examen común para todos los estados (*El País*, 16 de Marzo de 2010, p. 35). En el mismo sentido Wrigley (2007, p.

11) informa de cómo en Inglaterra, en los centros escolares de más bajo rendimiento, se han sustituido los exámenes normales en las diversas materias por otros más sencillos que “resultan muy útiles en las estadísticas gubernamentales, pero no ayudan a los alumnos a obtener un buen empleo o a entrar en la Universidad” (ibíd.). Esta misma tendencia puede observarse en las pruebas que se vienen realizando en Andalucía.

Las evaluaciones y pruebas de diagnóstico y la profesión docente

En lo que respecta a los docentes, se comprueba que las dificultades para obtener los resultados demandados, produce desánimo en el profesorado. Es cierto que en muchos centros difíciles se alcanzan buenos resultados, pero es difícil mantenerlos en el tiempo y, sobre todo, requieren un sobreesfuerzo que generalmente no se ve reconocido. Lo que ocurre entonces es la desafección y el abandono de la profesión:

El impacto más visible de esta cultura del rendimiento de alta presión ha sido una crisis importante de contratación, tal que ahora casi la mitad de los nuevos docentes que obtienen el título en Inglaterra abandonan en los dos primeros años y, en general, uno de cada diez profesores cada año. La crisis ha golpeado especialmente a aquellas escuelas que han tenido siempre más dificultades para atraer profesores (Wrigley, 2007, p. 49).

Algunos autores han señalado que las evaluaciones y pruebas de diagnóstico entrañan fenómenos de desmotivación profesional, “en ciertos casos, la percepción de la profesión se vuelve negativa, disminuye la satisfacción en la actividad profesional mientras que aparece la percepción de un nuevo estrés” (Mons, 2009, p. 124; traducción propia). Maroy y Cattonar (2002) analizan el rechazo de los docentes a determinadas formas de rendición de cuentas vinculadas a pruebas externas. Consideran, junto a otros autores (por ejemplo, Mons, 2009; Parveva, De Coster y Noorani, 2009) que ese rechazo no puede ser interpretado en clave meramente corporativa. A su juicio, las evaluaciones externas basadas fundamen-

talmente en el rendimiento de los alumnos generan problemas de identidad en la profesión docente. Como es sabido, la docencia se ha definido como una semi-profesión, ya que si bien la actividad del profesor está restringida por múltiples condicionantes, dispone de un relativo margen de actuación. Precisamente las evaluaciones basadas en pruebas estandarizadas cuestionan ese estatus, pues el profesor no será considerado como el único actor capaz de establecer un juicio definitivo y trascendente sobre las competencias adquiridas por sus alumnos, produciendo, por tanto, una descualificación de la profesión y del docente, a quien se le sitúa en una posición marginal frente a agentes externos.

¿Más riesgos que ventajas?

En fin, en un breve artículo publicado recientemente (Santos, 2010), Miguel Ángel Santos Guerra advierte de los riesgos, de los efectos quizás no deseados y, en suma, de los peligros que trae consigo el mecanismo de las pruebas de diagnóstico. Resumiendo su contenido, el profesor malagueño señala, en primer lugar, el problema, ya suscitado por otros autores, de que las evaluaciones de diagnóstico acaben por relegar a un segundo plano las áreas del currículo que no son objeto de evaluación pero que son igualmente importantes para la formación de los jóvenes. En segundo lugar, señala el peligro de que la evaluación se oriente exclusivamente a los resultados y no a los procesos, de manera que su última finalidad sea simplemente obtener buenos resultados en las pruebas, por lo cual se pregunta: “¿No bastaría para ello calificar a todos los alumnos y alumnas con un sobresaliente?” (Santos, 2010, p. 91). Se refiere, en tercer lugar, al problema que plantea el hecho de que las pruebas sean estandarizadas, lo que induce a comparar realidades que son incomparables. Finalmente, Santos Guerra nos advierte también del riesgo de que, en definitiva, la evaluación sólo exija cambios al profesorado, convirtiéndose, como se ha dicho en páginas anteriores, en un instrumento que más que ayudar a los docentes se utilice contra ellos.

Conclusión

Respecto a las evaluaciones y pruebas de diagnóstico, tan en boga en el discurso y prácticas dominantes en la política educativa actual, puede decirse aquello de que las cosas no son lo que parecen. Sin negar que este tipo de recurso pueda tener algún interés para producir conocimiento y para la mejora de la educación, es evidente que su potencialidad es mucho más limitada de lo que se pregona. En realidad se trata de un elemento más del dispositivo que conservadores y social-liberales aplican para el control de centros y profesores y para devaluar el sentido de la educación, orientándola básicamente hacia la mera consecución de unos resultados. Herederos, por tanto, de la vieja idea del *capital humano*, ese dispositivo funciona como un mando a distancia, pretendiendo gobernar desde fuera de los centros escolares lo que ocurre dentro de ellos. La rendición de cuentas, la evaluación y autoevaluación, los incentivos, el gerencialismo... y los resultados de las pruebas de diagnóstico constituyen los aspectos más visible de la nueva política. Se trata, por una parte, de un modo de gestión de la educación inspirada en el fordismo más clásico y, por otra, de una forma de trasladar la responsabilidad del fracaso a los centros y a los profesores.

Este dispositivo redefine el sentido de la educación en torno al saber hacer y en torno a los resultados, relegando a un segundo plano el conocimiento y los procesos. Sin embargo, a la vista, precisamente de los resultados, no parece que el mecanismo funcione con éxito, pues no sólo no logra avances significativos, sino que aumenta las diferencias entre distintos tipos de alumnos, y todo ello produciendo, además, una serie de "daños colaterales" que menoscaban el ya de por sí tímido efecto compensatorio de la escuela.

Aquí no se propone eliminar las evaluaciones y pruebas de diagnóstico, pero sí ponerlas en su lugar, desvelando que tras ellas hay un programa educativo oculto. El concepto y la práctica de la evaluación son susceptibles de diversas lecturas y tienen formas distintas de aplicación; no existe, como

se pretende, un único modelo. Por lo demás, conviene recordar que la educación es un sistema complejo, de manera que las viejas recetasseudocientíficas y tecnoburocráticas, son *mercancía averiada*, trasunto en realidad de una estrategia de control.

REFERENCIAS

- BARRENECHEA, I. (2010). Evaluaciones estandarizadas: seis reflexiones críticas. *Archivos Analíticos de Políticas Educativas*, 18 (8). En: <http://epaa.asu.edu/ojs/article/751> (consultado el 7 de mayo de 2010).
- BENÍTEZ HERRERA, A.J. (2010). Evaluar competencias básicas: ¿Una moda o una necesidad? *Cuadernos de Pedagogía*, 397, 82-85.
- BOLÍVAR, A. (1994). Autoevaluación institucional para la mejora interna. En: Zabalza, M.A. *Reforma educativa y organización escolar*. Santiago de Compostela: Tórculo, pp. 915-944.
- BOLÍVAR, A. (2001). Los centros educativos como organizaciones que aprenden: una mirada crítica. *Contexto educativo*, 18. En: <http://contexto-educativo.com.ar/2001/4/nota-04.htm> (consultado el 23 de octubre de 2009).
- CARABANA, J. (2009). *Las diferencias entre países y regiones en las pruebas PISA*. Documento 2 del Colegio Libre de Eméritos. En: http://www.colegiodeemeritos.es/docs/repositorio//es_ES//documentos/pisa_carabana_%28vf%29.pdf (consultado el 10 de Mayo de 2010).
- CRAHAY, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie*, 154, 97-110.
- EL PAÍS (2010): "Obama diseña un estándar educativo para todo EEUU". *EL PAÍS*, 16 de Marzo de 2010, p. 35.
- EVALUACIÓN DE DIAGNÓSTICO CURSO 2006-2007. Consejería de Educación. Junta de Andalucía. En: http://www.juntadeandalucia.es/educacion/portal/com/bin/Contenidos/OEE/evaluacion/informe_evaluacion_diagnostico_2006_2007/1181722468254_informe_finalx_12_de_junio.pdf (consultado el 12 de noviembre de 2009).
- EVALUACIÓN DE DIAGNÓSTICO CURSO 2007-2008. AVANCE. Consejería de Educación. Junta de Andalucía. En: [- 78 -](http://www.jun-</p>
</div>
<div data-bbox=)

- tadeandalucia.es/educacion/portal/com/bin/Contenidos/OEE/AvanceDiagnostico0708/1217447709315_avanceevaluaciondediagnostico0708.pdf (consultado el 12 de noviembre de 2009).
- EVALUACIÓN DE DIAGNÓSTICO CURSO 2008-2009. Consejería de Educación. Junta de Andalucía. En: http://www.juntadeandalucia.es/educacion/portal/com/bin/Contenidos/OEE/evaluacion/Evaluaciondiagnostico0809/1246617499991_evaluaciondiagnostico2008_09.pdf (consultado el 12 de noviembre de 2009).
- FERRAO, M^ªE. (2009). Sensibilidad de las especificaciones del modelo valor añadido: midiendo el estatus socioeconómico. *Revista de Educación*, 348, 137-152.
- GARCÍA PÉREZ, J.I., PINTO PRADAS, J.L. y ROBLES ZURITA, J.A. (2009). *Diferencias regionales y rendimiento educativo. Análisis de resultados del estudio PISA-2006*. Sevilla: Centro de Estudios Andaluces.
- GIL FLORES, J. (2009). Hábitos y actitudes de las familias hacia la lectura y competencias básicas del alumnado. *Revista de Educación*, 350, 301-322.
- GIMENO SACRISTÁN, J. (comp.) (2008). *Educación por competencias, ¿qué hay de nuevo?* Madrid: Morata.
- GOLDSTEIN, H. (2008). Comment peut-on utiliser les études comparatives internationales pour doter les politiques éducatives d'informations fiables? *Revue française de pédagogie* 164, juillet-août-septembre 2008, 69-76.
- GRANT, S.G. (2000). Teachers and Tests: Exploring Teachers' Perceptions of Changes in the New York State Testing Program Education. *Policy Analysis Archives* 8, (14). En: <http://epaa.asu.edu/epaa/v8n14.html> (consultado el 2 de diciembre de 2009).
- JONES, L. y MOORE, R. (2008). Apropiación del concepto de competencia: competencias, movimiento de la Nueva Derecha y el proyectado cambio cultural. *Profesorado. Revista de Currículum y Formación del Profesorado*, 12 (3). En: <http://www.ugr.es/local/recfpro/rev123ART7.pdf> (consultado el 18 de noviembre de 2009).
- LÓPEZ MARTÍN, E. y otros (2009). Estudio de variables determinantes de eficiencia a través de los modelos jerárquicos lineales en la evaluación PISA 2006: el caso de España. *Archivos Analíticos de Políticas Educativas*, 17 (16). En: <http://epaa.asu.edu/ojs/article/viewFile/18/18> (consultado el 4 de enero de 2010).
- MARROY, C. y CATTONAR, B. (2000). Professionnalisation ou désprofessionnalisation des enseignants? Le cas de la communauté française de Belgique. *Cahiers de recherche en éducation et formation*, 18.
- MARIO, C. y MANGEZ, C. (2008). Rationalisation de l'action publique ou politisation de la connaissance? *Revue française de pédagogie*, 164, juillet-août-septembre 2008, 87-90.
- MINISTERIO DE EDUCACIÓN Y CIENCIA. SECRETARÍA GENERAL DE EDUCACIÓN. INSTITUTO DE EVALUACIÓN (2007). *PISA 2006. Programa para la Evaluación Internacional de Alumnos de la OCDE. Informe español*. Madrid: MEC.
- MONS, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie*, 169, 99-140. Disponible también en: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111FR.pdf
- NOTA DE PRENSA (18 de octubre de 2006). Consejería de Educación. En: http://www.juntadeandalucia.es/educacion/nav/contenido.jsp?pag=/Contenidos/Gabinete-Prensa/Notas_de_prensa/2006/Octubre/nota_prensa_181006_pruebadiagnostico_parlamento&vismenu (consultado el 18 de marzo de 2010)
- OCDE (2008). *Competencias científicas para el mundo de mañana*. Santillana, e-books: <http://browse.oecdbookshop.org/oecd/pdfs/browseit/9807014E.pdf>.
- PADILLA, R.A. (2009). Exámenes internacionales y nacionales. Encuentros y desencuentros. *Perfiles Educativos*, Tercera época, XXXI (123), 44-59. En: <http://www.iisue.unam.mx/seccion/perfiles/> (consultado el 10 de Mayo de 2010).
- PARVEVA, T. (coord.), DE COSTER, I. y NOORANI, S. (2009). *Les évaluations standardisées des élèves en Europe: objectifs, organisation et utilisation des résultats*. Bruxelles: Agence exécutive "Éducation, audiovisuel et cultura", P9 Eurydice. En: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109FR.pdf (consultado el 17 de Mayo de 2010).
- ROCHEX, J-Y. (2008). *PISA et les enquêtes internationales*. Enjeux scientifiques, enjeux poli-

- tiques. *Revue française de pédagogie*, 164, juillet-août-septembre 2008, 81-85.
- SANTOS GUERRA, M.A. (2010). Los peligros de la evaluación. *Cuadernos de Pedagogía*, 397, 90-93.
- SHARON, N. y BERLINER, D. (2007). *Collateral Damage. How Highstakes testing corrupts America's schools*. Cambridge, Mass: Harvard Education Press.
- SLEE, R., G. y otros. (2001). *¿Eficacia para quién? crítica de los movimientos de las escuelas eficaces y de la mejora escolar*. Madrid: Akal.
- STEINBERG, J. (2000). Student Failure Causes States To Retool Testing Programs. *New York Times* 22-12-2000. En: <http://www.nytimes.com/2000/12/22/us/student-failure-causes-states-to-retool-testing-programs.html?scp=3&sq=Steinberg&st=nyt> (consultado el 12 de septiembre de 2007).
- TRÖHLER, D. (2009). Armonizar el globo educativo. Política mundial, rasgos culturales y los desafíos a la investigación educativa. *Profesorado. Revista de currículo y formación del profesorado*, 13 (2). En: <http://www.ugr.es/~recfpro/rev132FIRINV1.pdf> (consultado el 10 de Mayo de 2010).
- WRIGLEY, T. (2007). *Escuelas para la esperanza. Una nueva agenda hacia la renovación*. Madrid: Morata.