

¿QUÉ DICE LA INVESTIGACIÓN SOCIAL SOBRE LAS PRUEBAS EDUCATIVAS EXTERNAS?

RAFAEL FEITO ALONSO

Universidad Complutense de Madrid

rfeito@cps.ucm.es

ORCID iD: <http://orcid.org/0000-0002-2579-5028>

Cómo citar este artículo / Citation: Feito Alonso, R. 2017. "¿Qué dice la investigación social sobre las pruebas educativas externas?". *Revista Internacional de Sociología* 75 (1): e053. doi: <http://dx.doi.org/10.3989/ris.2017.75.1.15.89>

WHAT DOES SOCIAL RESEARCH SAY ABOUT HIGH-STAKES TESTS?

Copyright: © 2017 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia *Creative Commons Attribution (CC BY)* España 3.0.

Recibido: 09/06/2015. **Aceptado:** 13/05/2016.

Publicado on line: 02/03/2017

RESUMEN

Las pruebas externas estandarizadas (*high-stakes tests*, en la jerga anglosajona), que han de ser aprobadas para obtener una credencial de Educación Secundaria, han dado lugar a amplios debates en los países en que se han implantado. Especialmente en el caso de los Estados Unidos de América se ha detectado que este tipo de test ha provocado una reducción considerable de los contenidos curriculares, los cuales se han centrado en exceso en lo que se pueda preguntar en los propios test. Al mismo tiempo, se ha producido una modificación sustancial en los procesos de aprendizaje que, centrados en unos conocimientos de tipo fáctico, se alejan de la experimentación, el debate o el trabajo en equipo. Pese a que hay discrepancias, en general, la investigación ha detectado que es dudoso que estas pruebas mejoren los resultados académicos. Sin embargo, se incrementa el porcentaje de abandono escolar temprano. Ha sido frecuente cierto nivel de fraude en las pruebas por parte de los directivos escolares, evitando que se presentasen a los test los alumnos de menor rendimiento académico. La revisión de investigaciones que se lleva a cabo en este artículo permite contextualizar el debate que en España se ha generado en torno a las pruebas externas que plantea la Ley Orgánica de Mejora de la Calidad Educativa (LOMCE) aprobada en 2013.

PALABRAS CLAVE

Educación comparada; *High-stakes*; Legislación escolar; LOMCE; Pruebas externas; Reforma educativa.

ABSTRACT

High-stake tests, which students need to pass in order to gain a Secondary education certificate, have aroused a lot of controversy whenever they have been implemented. Especially in the USA these tests have produced a dramatic shrinking of school knowledge as they have been focused into questions posed by the very tests. At the same time, there has been a critical modification of the learning processes due to the fact that these tests encourage students to pay more attention to factual knowledge, which is far away from hands-on learning, debating in class or working in teams. In spite of certain discrepancies, by and large, research casts serious doubts about whether these tests are conducive to better academic performance. Nevertheless, the dropout rate is on the rise. More often than not, school principals have been found preventing struggling students to sit the exams. A review of the research into this matter allows us to put into context the debate surrounding external exams posed by the current education law (LOMCE: Law for the Improvement of Educational Quality) passed by the parliamentary majority of the Popular Party in Spain in 2013.

KEYWORDS

Comparative Education; Educational Reform; External Tests; High-Stakes; LOMCE; School Law.

La LOMCE (Ley Orgánica de Mejora de Calidad de la Educación), la vigente ley educativa aprobada por el parlamento español a finales de 2013, propone la realización de pruebas externas para todo el alumnado de Primaria (en los cursos tercero y sexto), de la Secundaria obligatoria y del Bachillerato. En estos dos últimos niveles sería preciso aprobarlas para obtener la credencial correspondiente. Esto es lo que en la jerga anglosajona se llaman exámenes de tipo *high-stakes*, expresión que quiere decir que se trata de pruebas que tienen consecuencias importantes para quien las lleva a cabo. Un ejemplo de *high-stake*, fuera del ámbito escolar, es el examen de conducir: quien no lo aprueba no está autorizado a circular con un automóvil por las vías públicas. También existen los *low-stakes*, exámenes –como, por ejemplo, los que realiza el PISA– que no inciden directamente sobre el futuro del alumnado –nada tienen que ver con su promoción académica–, del profesorado o de los centros escolares. Otro ejemplo de prueba *low-stake* es la finlandesa *Ylioppilastutkinto* (cuyo precedente es el examen de acceso a la Universidad de Helsinki), que es el único examen nacional en este país y sirve como estándar de comparación para valorar el desempeño de los estudiantes de modo que las escuelas puedan elaborar estrategias de mejora (Klein 2010). En alguna ocasión, las pruebas son tan sumamente exigentes (*high*) que el fracaso de los alumnos en ellas puede llegar a suponer el cierre de la escuela y el despido de sus profesores. De este modo, en el caso de los Estados Unidos, la ley *No Child Left Behind* (NLCB)¹ determina que, en el segundo año en que una escuela tenga malos resultados, a su alumnado se le ofrecerá la posibilidad de acudir a un centro de mayor calidad. De no mejorar en el tercer año, sus estudiantes pueden hacer uso de tutorías gratuitas después del horario escolar. Si la escuela siguiera con una baja puntuación, pasaría a convertirse en un centro concertado (*charter school*) y su plantilla disminuiría drásticamente.

Los resultados de estas pruebas pueden ir más allá de lo estrictamente escolar. En los Estados Unidos sirven para configurar zonas de escolarización y, en consecuencia, escuelas socialmente excluyentes (Amrein y Berliner 2002). Las agencias de propiedad inmobiliaria utilizan tales resultados para valorar la calidad del vecindario y esto afecta al precio de la vivienda. Las familias que pueden permitírselo deciden vivir en una zona u otra en función de estos resultados.

Los grupos políticos, sociales y profesionales opuestos a la LOMCE consideran que las pruebas externas propuestas en esta ley son una reedición de las reválidas que existieron antes de la aprobación de la Ley General de Educación de 1970 –cosa que el ministro de Educación que promovió la ley, José Ignacio Wert, negaba² alegando que estas eran un sistema de *numerus clausus* que regulaba el nú-

mero de personas que podían pasar de la Primaria al Bachiller Elemental y de este al Superior–. Es decir, de acuerdo con quienes se oponen a la ley, se trataría de un mecanismo de selección socioeducativa.

En principio, poco habría que objetar a la existencia de pruebas que permitan unificar los criterios de rendimiento de los estudiantes. El mayor problema que plantean, salvo que sean particularmente abiertas y flexibles, es que tienden a condicionar la pedagogía y los contenidos de los cursos previos. Del mismo modo que el segundo curso de Bachillerato se ha convertido en una especie de academia para la selectividad, cuarto de la ESO podría experimentar una transformación similar debido a las pruebas externas.

Es muy legítimo que tanto al Gobierno como a la ciudadanía, a las familias y al profesorado les preocupe saber con cierto grado de precisión si nuestros escolares han aprendido aquello que es exigible al finalizar determinado curso o al transitar de una etapa educativa a otra. Diane Ravith, inicialmente una defensora de este tipo de pruebas, lo explicaba con claridad meridiana:

“¿Quién podría oponerse al hecho de asegurar que los niños dominan las destrezas básicas de la lectura y de las matemáticas? ¿Quién podría oponerse a un examen anual de estas destrezas? Con seguridad, yo no. ¿No examinan, de hecho, las escuelas a sus alumnos al menos una vez al año?” (2010: 25).

En el caso de España, en ausencia de pruebas externas, el alumnado puede recorrer toda la enseñanza preuniversitaria –hasta la realización de la Prueba de Acceso a la Universidad (PAU)– con la única y exclusiva evaluación de los profesores de su centro (el cual podría ser solo uno en el supuesto de haber estudiado en un centro privado o, un mínimo de dos, en el caso de haberlo hecho en la enseñanza pública). Todo esto es un terreno abonado para la arbitrariedad: se podría aprobar o suspender una asignatura en función de las peculiaridades de su profesor. Lo que tenemos ahora en nuestro sistema educativo supone que no sabemos qué pueda haber aprendido un alumno. Si se presentara a la PAU, algo –aunque no mucho– se podría saber. Las pruebas externas se convertirían en una garantía de que el alumnado que las aprobase ha adquirido al menos los conocimientos sobre los que versaren tales evaluaciones.

No obstante, conviene tener en cuenta el peligro que supone magnificar el peso de estas pruebas en la evaluación del sistema educativo. Hace ya algunas décadas, Donald T. Campbell (1976) acuñó una famosa ley que lleva su nombre (también conocida como el “principio de indeterminación de Heisenberg para las ciencias sociales”) que viene a decir que cuanto más se utiliza un indicador social para resolver problemas sociales, resulta más susceptible de distorsionarse y de corromper los procesos sociales que trata de corregir. En su libro sobre los exámenes del tipo *high-stakes*, Nichols y Berliner (2008a) ad-

vertían de la similitud en la práctica de estas pruebas con el uso perverso que hizo Sears en California en su retribución a los mecánicos de coches. En lugar de abonarles un salario, la empresa decidió que se les pagaría una comisión en función del número de reparaciones que hicieran. Al poco tiempo, las autoridades del estado amenazaron a Sears debido a la enorme cantidad de quejas de clientes por reparaciones innecesarias. Algo similar ocurrió durante años en la Unión Soviética cuando se ordenaba a los trabajadores fabricar, bajo amenaza, tractores o tanques, con independencia de que posteriormente fueran útiles o no.

La bibliografía sobre pruebas externas es inmensa. Para seleccionar artículos científicos se ha recurrido a bases de datos como *Education Resources Information Center*, *JSTOR* y *Sociological Abstracts* utilizando palabras clave del tipo “*high-stakes exams*”, “*exit exams*” o “*exit test*”. Ha sido de gran utilidad el uso del buscador general de Google, ya que esto ha permitido acceder no solo a artículos científicos, sino también a publicaciones de la prensa general o del ámbito profesional docente, y a informes realizados por diversos *think tanks*. A lo largo de la revisión bibliográfica se ha producido una suerte de efecto de bola de nieve: muchas de las lecturas remitían a otras. Esto ha permitido conocer en profundidad las réplicas, contrarréplicas, dúplicas y demás a que ha dado lugar este intenso debate protagonizado fundamentalmente por economistas, sociólogos, psicólogos e historiadores. Buena parte de la literatura sobre este tema procede de la discusión generada en torno al uso de los *high-stakes tests* en los Estados Unidos de América.

En las páginas que siguen se explicarán el debate sobre estas pruebas y las consecuencias que han tenido; las cuales se refieren; fundamentalmente, a la reducción de los contenidos curriculares y al modo en que se enseña, al desempeño educativo, a los diferentes resultados obtenidos en función del estatus socioeconómico y de la etnia del alumnado, y, finalmente, a las repercusiones que han tenido sobre las actitudes de los equipos directivos, de los profesores y de los estudiantes. Todo ello podría servir de orientación para el debate educativo que ha suscitado en España la LOMCE. La experiencia acumulada sobre este tipo de pruebas en otros países debería hacernos muy cautos, no ya tanto con respecto a su existencia, como con relación a su contenido y al tipo de preguntas que en ellas se planteen.

¿QUÉ CONTENIDOS EVALÚAN LAS PRUEBAS EXTERNAS?

Quizás esta sea la cuestión principal del debate relativo a las pruebas externas. Como se verá más adelante, estas no tienen por qué reducir los contenidos curriculares, ni centrarse en unas pocas ma-

terias del currículo escolar, ni tienen que consistir en preguntas de tipo test. De hecho, dos autores extremadamente críticos con los *high-stakes*, como Nichols y Berliner (2008a), no se oponen a los exámenes estandarizados y no tienen problema alguno en reconocer que el SAT (*Scholastic Aptitude Test*, prueba similar a la de acceso a la universidad en España) puede ser una prueba de gran utilidad para la medición de saberes complejos y diversos.

El principal riesgo de este tipo de pruebas es que el conjunto de la enseñanza se pueda concentrar en los contenidos que en ellas se plantean (*teaching to the test*) a costa de otros igualmente necesarios para la formación de las nuevas generaciones. Un trabajo realizado por J. J. Pedulla (citado en Nichols y Berliner 2008a) indica que el 80% de los profesores de aquellos estados norteamericanos en los que hay *high-stakes* se queja de que son presionados para obtener buenos resultados, lo que se traduce en enseñar para los test obviando todo lo demás.

Los test que se han venido utilizando limitan considerablemente el aprendizaje de destrezas indispensables para desenvolverse en el mundo actual. Se trata de una estrategia que, a corto plazo, puede beneficiar a los alumnos de las clases media y alta, pero que resulta contraria a lo que les va a esperar en el mundo del trabajo. Tal limitación anula la posibilidad de centrarse en la profundización de los temas curriculares, en desarrollar niveles sofisticados de aprendizaje, los cuales incluyen destrezas de alto nivel y la capacidad para aplicarlas en escenarios que no resulten familiares al alumnado. Esto es lo que al respecto decía Robert Reich (quien fuera secretario de Trabajo en la primera administración Clinton):

“Muchos de los nuevos empleos dependen de la creatividad –del pensamiento *out of the box*–, de la originalidad y del talento. Casi por definición, los test estandarizados no pueden medir este tipo de cosas. Son adecuados para medir la habilidad para repetir mecánicamente hechos y aplicar modos estandarizados de análisis. Nuestra obsesión por los test estandarizados va justamente en dirección contraria a las necesidades de la economía moderna” (Reich 2001: 64).

La investigación es concluyente con respecto a la reducción del conocimiento a lo meramente fáctico. Au (2008) señala que la enseñanza de la historia –cuando esta materia entra en los test– se convierte en una colección de datos fácticos, lo que impide que sea una actividad que promueva el pensamiento en torno a esta disciplina. Richard Rothstein (*cf.* Nichols y Berliner 2006) explica que en los test se pide a los estudiantes que identifiquen detalles de los textos, en lugar de centrarse en la idea principal. Por ejemplo, un escrito sobre Cristóbal Colón puede pedir a los alumnos que citen los nombres de las tres carabelas, de modo que estos no lleguen a saber que al navegar hacia el oeste el

navegante genovés quería probar que la tierra era redonda. En el mismo sentido, Anagnostopoulos (2003) observó que los test en Chicago limitaban la posible interpretación global de un texto como *Matar a un ruiseñor*, excluyendo elementos fundamentales de naturaleza política e ideológica. Se corre el serio riesgo de que la docencia se convierta en una suerte de edición del juego del *Trivial Pursuit*. Hamilton (2005) señala que en los Estados Unidos la instrucción empieza a parecerse a los test. Por ejemplo, las actividades derivadas de la lectura de un libro se reducen a leer pasajes cortos seguidos de preguntas de elección múltiple, un tipo de lectura –si es que así cupiera llamarla– que no se da en la vida real.

Las pruebas estandarizadas suelen abordar cuestiones de unas pocas materias del currículo, aquellas que la escuela considera importantes, dejando el resto en la penumbra de la indiferencia. Este sería el caso de las pruebas de “Conocimientos y Destrezas Indispensables” que realiza la Comunidad de Madrid a alumnos de sexto de Primaria y tercero de la ESO (se puede ver en qué consisten en la web de la Consejería de Educación de esta comunidad autónoma). Se trata de pruebas que se centran de modo casi exclusivo en la competencia lectora y en conocimientos matemáticos obviando el resto de contenidos curriculares; es decir, tan solo analizarían el grado de adquisición de ciertas competencias básicas. Igualmente se ha denunciado el elevado coste de la realización de estas pruebas.

En el caso de la LOMCE nos encontramos con una situación similar. Así, las pruebas de tercer curso de Primaria se limitan a tan solo dos de las siete competencias contempladas en el Real Decreto 126/2014, de 28 de febrero (el cual determina el currículo básico de Primaria). Sin que medie explicación alguna, las pruebas externas de este curso se limitan a tan solo la Competencia Lingüística y a la Competencia Matemática (aunque, en realidad, la segunda de las competencias se llama “Competencia Matemática y Competencias básicas en Ciencia y Tecnología”).

De este modo, estas pruebas no abordan las siguientes competencias (ordenadas desde la tercera a la séptima):

- 3.º Competencia Digital.
- 4.º Aprender a Aprender.
- 5.º Competencias Sociales y Cívicas.
- 6.º Sentido de Iniciativa y Espíritu Emprendedor.
- 7.º Conciencia y Expresiones Culturales.

Finalmente, conviene poner de manifiesto cómo el nivel de exigencia de las pruebas varía a lo largo del tiempo. Este ha sido el caso de las pruebas de sexto de Primaria de la Comunidad de Madrid.

De un curso a otro, los resultados pueden mejorar notablemente. Pese a que no cabría descartar un perfeccionamiento del aprendizaje de los alumnos, la clave para tal evolución parece residir en que el examen es más fácil. Sin ánimo de exhaustividad, y a modo de ejemplo, el dictado pasa de ser un texto –uno o varios párrafos– a un listado de pequeñas frases. Volviendo a los Estados Unidos, el caso de Chicago resulta ejemplar por el modo en que en cuestión de muy poco tiempo se puede pasar de obtener pésimos resultados a mejorar considerablemente, bien como consecuencia de que profesorado y alumnado aprenden a desenvolverse con estos test, o bien de que simplemente se rebajan sus niveles de exigencia (Roderick y Nagaoka 2005). En definitiva, su rigor depende en muy buena medida de la propia severidad de las pruebas. En los Estados Unidos fluctúa considerablemente entre estados, y en cada uno de ellos, de un año a otro. Un informe de la Fundación Abell (2003) hacía referencia al dilema que tenía que afrontar la Junta Estatal de Educación de Maryland sobre el rigor de las pruebas. De llegar a la misma exigencia que la media del resto de los estados, se hubiera condenado a la no obtención del título de Secundaria Superior a un 20% más de estudiantes que en el curso anterior, con el agravante de que hubiese afectado sobre todo a los de menor nivel económico.

¿CÓMO EVALÚAN LAS PRUEBAS EXTERNAS?

No menos importante que los contenidos de las pruebas externas, es el modo cómo estas evalúan. Hasta ahora, lo que ha predominado es el examen tipo test, algo que llevó a que, durante su primera campaña presidencial, Obama apuntase la necesidad de acometer cambios sustantivos en la ley NCLB (que ampara, pese a que no obliga a ello, los exámenes externos). Llegó a decir que los profesores dejarían de estar “abocados a dedicar el curso a preparar a los niños para rellenar círculos en test estandarizados”. Señaló que materias como la historia y el arte estaban siendo marginadas y que, en definitiva, niños y adolescentes no recibían una educación completa. Propuso que era preferible ayudar a las escuelas con problemas en lugar de sancionarlas (Ravitch 2009). Sin embargo, finalmente los exámenes externos tipo test continúan existiendo.

Las pruebas externas definen los contenidos que cuentan como conocimiento legítimo, de modo que lo que no se evalúa en los test –lo que habitualmente sucede con el arte, las ciencias y los estudios sociales– mengua considerablemente (Jones, Jones y Hargrove 2003; Au 2008; Mons 2009;). Además, y esto es lo que se quiere plantear en este epígrafe, los test terminan por controlar el modo en que se enseña, de manera que el conocimiento se plantea en forma de preguntas de respuesta corta, precisa e in-

discutible. A partir de aquí, el currículo es presentado como un conjunto de datos aislados que los alumnos han de memorizar. Los profesores asumen un mayor control sobre la didáctica, de modo que los alumnos tienen menos oportunidades de dirigir su propio aprendizaje o de explorar conceptos o temáticas que pudieran interesarles.

El peligro de que la formación se centre en los test es evidente. Jonathan King (citado en Wagner 2010), un reconocido biólogo molecular del MIT, envió a sus dos hijos a la misma escuela. Su hijo mayor aprendió desde la experiencia. Junto con sus compañeros de clase, iba a una charca cercana a su centro escolar y allí tomaba muestras. De hecho, los niños descubrían criaturas que la ciencia desconocía. Con su segundo hijo, las cosas fueron radicalmente diferentes ya que había sido preparado para pasar los test. La experiencia, el mancharse las manos, habían desaparecido. Howard Gardner (2002) plantea la necesidad de que el alumnado piense y actúe como lo hacen los científicos, los matemáticos, los historiadores y los artistas. Las investigaciones han demostrado que los alumnos que aprenden de este modo manifiestan una elevada motivación intrínseca y están más dispuestos a desarrollar el pensamiento crítico. Justamente, esto es lo que se propone el interesantísimo informe Grattan:

“Las reformas acometidas en Hong Kong giran en torno a un objetivo central: mejorar el aprendizaje de los estudiantes. Un análisis en profundidad detectó que el aprendizaje de los estudiantes era monótono y orientado hacia el examen y dejaba poco espacio para pensar, investigar y crear. La enseñanza se había convertido en un proceso de transmisión unidireccional. Hong Kong quería que los niños se implicaran en actividades de aprendizaje a partir de lo que saben, interactuando, creando e investigando nuevos conocimientos. Su nueva definición de aprendizaje supuso alejarse del énfasis previo en la adquisición de conocimiento” (Grattan 2012: 18).

Las pruebas externas de Secundaria³ que propone la LOMCE pretenden, en principio, evaluar una serie de destrezas cognitivas de alto nivel. Sin ánimo de ser exhaustivo, he aquí –entre otras cosas– lo que se quiere evaluar de la Competencia en Comunicación Lingüística:

- a) Inferir la información relevante de los textos, identificando la idea principal y las ideas secundarias, y estableciendo relaciones entre ellas.
- b) Reconocer la coherencia global del texto.
- c) Resumir textos narrativos, descriptivos, expositivos y argumentativos de forma clara, integrando la información en oraciones que se relacionen lógicamente y semánticamente.
- d) Deducir información y relaciones no explícitas así como organizar la información del texto para vincularla a un conocimiento previo.

O, en el caso de la Competencia Matemática y de la Competencias Básicas en Ciencia y Tecnología se evaluarían, entre otras, estas capacidades:

- a) Determinar métodos, estrategias o herramientas eficientes y apropiados para resolver problemas para los que hay métodos de resolución comúnmente utilizados.
- b) Utilizar modelos: generar ecuaciones o diagramas para representar o ilustrar situaciones, sistemas, procesos, conceptos, principios o relaciones.

Sin embargo, al final del documento oficial que explica el contenido de estas pruebas, el lector descubre que se trata de exámenes tipo test (“Las preguntas de la prueba serán en formato de elección múltiple con cuatro alternativas y de respuestas semiconstruidas”) y que contendrán 350 cuestiones. Parece poco probable que se puedan generar ecuaciones o reconocer la coherencia global de un texto en un test. Daría la sensación de que no se hubieran puesto de acuerdo entre sí los autores de las páginas referidas a qué se quiere evaluar con los de las relativas a cómo evaluar.

La medición de saberes complejos tiene un coste económico sensiblemente más alto –en torno al triple– que el de los contenidos, relativamente simples, de un test. Así, por ejemplo, el estado norteamericano de Maryland contrató a una empresa privada – *Measurement Inc.*, como se puede ver en Abell Foundation (2003)–. Sin embargo, las respuestas abiertas y los ensayos eran evaluados siguiendo el criterio de la evaluación por pares realizada por profesores (en caso de que hubiese gran desacuerdo entre ellos, la evaluación se asignaría a un tercer evaluador).

En tiempos pasados, los test iban más allá de lo meramente fáctico. Desde los años sesenta hasta los noventa del pasado siglo, la evaluación en los Estados Unidos se basaba en la *National Assessment of Educational Progress* (NAEP). Inicialmente pensada como un mecanismo de información basado en test psicométricos, la NAEP fue evolucionando hacia un sistema que incluía la medición de una amplia gama de contenidos cognitivos y no cognitivos, destrezas y rasgos del carácter personal. También se valoraban los hábitos de buena salud y la capacidad de cooperar con otras personas para resolver problemas (Ruiz i Tarragó 2010). Las evaluaciones incluían test, pruebas creativas y observaciones de las actividades de los alumnos y de los centros. A causa de su alto coste económico, las metodologías de evaluación de la NAEP fueron abandonadas durante los recortes presupuestarios de los años setenta y nunca más fueron recuperadas.

Por tanto, nada impide que las pruebas externas vayan más allá de los test y sean capaces de medir destrezas complejas. Justamente esto es lo que

ocurre en diversos países. Un examen de nivel que contemple destrezas genéricas tales como el modo en que se redacta, se razona, se procesa la información, etc. podría ser de gran utilidad. En concreto este sería, por ejemplo, el modelo de las *Coalition Schools* en los Estados Unidos, en las que el título de Educación Secundaria se obtiene en pruebas públicas, similares a las de nuestras tesis doctorales, ante tribunales constituidos por profesores del centro y por algún adulto nombrado por el propio estudiante (Apple y Beane 2005). Si se quisiera una evaluación más objetiva se podría contar con la presencia de profesores de otros centros. En este tipo de pruebas, cada estudiante presenta en público investigaciones o ensayos para cada una de las áreas de conocimiento, sean asignaturas o bloques curriculares más interdisciplinarios. De paso solventaríamos en buena medida el actual carácter escasamente democrático de la evaluación, al convertirla en una actividad pública y genuinamente colegiada. Luis Garicano (2015) explicaba una forma similar de examen de fin de etapa, esta vez en Holanda:

“Recientemente tuve la oportunidad de asistir [...] a la presentación de los proyectos de final de bachillerato en un Technasium. Un Technasium es un instituto de bachillerato especializado para estudiantes que quieren en el futuro estudiar materias técnicas y científicas, desde ingenierías a matemáticas o ciencias naturales. [...] En el Technasium los proyectos adquieren una importancia especial. El proyecto de fin de estudios (último año de bachillerato) requiere 200 horas de trabajo por estudiante (equivalente a cinco semanas de trabajo a tiempo completo).

En el Technasium de Amersfoort, observé fascinado la presentación de seis proyectos. Un grupo tenía que diseñar, bajo la supervisión de un despacho de arquitectura, la infraestructura de una pequeña urbanización de vacaciones: carreteras, energía sostenible, puentes. Los cálculos incluían el tipo de puentes sobre el canal y sus soportes, el grosor de las carreteras, el tipo de energía usada, la forma de guardar el exceso de energía. Otro grupo tenía el encargo de diseñar un sistema para ayudar a los ancianos a levantarse de la cama sin ayuda. Resolvieron el reto con la ayuda de un brazo articulado para la parte alta del cuerpo y un soporte con un pequeño motor para las piernas, todo ello controlado con un sencillo control remoto. Un tercer grupo investigó las anomalías cromosómicas en dos pacientes del Hospital Universitario de Utrecht. Un cuarto grupo diseñó un robot que pudiera llevar bebidas de una habitación a otra, evitando obstáculos. Otro grupo nos dejó con la boca abierta cuando tuvo que disculparse al explicar que su presentación había sido declarada confidencial por el cliente: el cliente, una empresa líder en tratamiento de aguas, había decidido que el proyecto había desarrollado conocimiento patentable y no quería que nada fuera presentado hasta que existiera la patente”.

En el caso de países como “Australia, Dinamarca, Inglaterra, Escocia, Finlandia, Francia, Irlanda,

Holanda y buena parte de Canadá y Alemania, por ejemplo, los exámenes de fin de la Secundaria superior [...] tienen lugar durante un periodo de dos semanas o más. Los exámenes de cada asignatura duran alrededor de tres horas y exigen que los estudiantes escriban ensayos, describan experimentos y muestren qué pasos han seguido para resolver un problema” (Bishop 2005: 4).

Tal y como explicaba Sireci (2005), un examen estandarizado significa que su contenido es equivalente en todos los lugares en que se aplica y que las condiciones a que se somete a los sujetos examinados son las mismas. A esto hay que añadir que una prueba estandarizada no tiene por qué limitarse tan solo a preguntas de respuesta múltiple –de tipo test– y que su contenido puede abarcar cuestiones de alto nivel cognitivo (en las que defender un punto de vista o desarrollar un argumento), para lo que aduce como ejemplo el *Graduate Records Exam* o el *Law School Admissions Test* (fácilmente consultables en Internet). Este también podría ser el caso de un examen de ámbito mundial de conocimientos de inglés como es el *Cambridge Proficiency in English* –igualmente de fácil acceso en la red–. Aquí (y solo apunto algunos aspectos), junto con una enorme batería de preguntas de respuesta múltiple y ejercicios en los que acertar con la palabra adecuada para completar una frase o escribirla de un modo alternativo, hay que redactar varios escritos en los que sopesar diferentes puntos de vista y realizar un ejercicio de conversación. En definitiva, la opción del examen tipo test es una de entre otras muchas posibilidades.

Cocker (2005) menciona la existencia de exámenes en los que se requiere escribir ensayos cortos donde se responde a preguntas del siguiente tenor: “explica por qué motivos has acudido al médico y describe los síntomas”, “describe cómo ha sucedido el accidente y, si hace falta, haz un diagrama”. En esta misma línea, Plake (2005) hacía referencia a los exámenes de desempeño en los que se exige acometer una tarea similar a la que se tiene que ejecutar en una situación real. Por ejemplo, para convertirse en dentista se han de demostrar destrezas con un paciente real.

RESULTADOS

Una de las cuestiones clave en este debate es el de si con estos exámenes se consigue elevar el nivel educativo del estudiantado. Los datos procedentes de las evaluaciones internacionales como el PISA o el TIMSS⁴ apuntan a la prudencia. Uno de los datos aportados desde los defensores de la LOMCE en favor de las pruebas externas es esta cita del informe de la OCDE *Education at a Glance. 2012*:

“Los datos del PISA de 2009 sugieren que los países que utilizan exámenes externos estandarizados tienden a obtener mejores resultados, incluso cuan-

do se considera la renta nacional. Más del 10% de la varianza en los resultados de lectura entre países se explican por la existencia de estos exámenes. En otras palabras, los estudiantes de sistemas escolares que usan exámenes externos estandarizados están 16 puntos por encima de media que los estudiantes de sistemas que no los utilizan [...]. ...la existencia de estos exámenes no está relacionada con la equidad, esto es la solidez de la relación entre el origen socioeconómico de los estudiantes y el desempeño en lectura. Estos hallazgos sugieren la necesidad de más investigaciones referidas a la relación entre la existencia de los exámenes estandarizados y el desempeño relativo de los estudiantes en evaluaciones como el PISA (OECD 2012: 527).

Como se puede ver, la última frase invita a la cautela. Algo similar ha sucedido con las pruebas del TIMSS. Con los datos de 39 países referidos al octavo curso de esta evaluación internacional en el año 1995, Bishop (2006) descubrió que aquellos países que contaban con sistemas de exámenes externos basados en el currículum (CBEEES: *Curriculum-based external exit examination system*) tienen mejores resultados en Matemáticas y Ciencias (habiendo controlado previamente las diferencias de renta per capita entre los países). Una investigación posterior, acometida por Huang (2009), reanaliza estos mismos datos, incluyendo en el listado de los 39 países de Bishop dos más excluidos por este autor que, en realidad, sí cuentan con CBEEES: Chipre y Portugal (no obstante, Huang comenta que podría haberse incluido también España si se considerase la PAU como un CBEEES). La mera inclusión de estos dos países hace que las diferencias de rendimiento detectadas por Bishop se conviertan en insignificantes. Huang considera que se deben básicamente al modo en que en cada nación se aplican esos exámenes. Recuérdese, tal y como se ha explicado más arriba, cómo son los exámenes externos en Holanda.

Si el objetivo del Ministerio de Educación español con las pruebas externas es mejorar en los informes PISA, se podría tomar en consideración el caso de los Estados Unidos. Pese a los ya numerosos años en que se llevan aplicándolas, este país no ha conseguido mejorar posiciones en las sucesivas ediciones de los informes PISA. Esto es lo que afirmaba la que quizás sea la investigación más concluyente, la llevada a cabo por el Consejo Nacional de Investigación (Hout y Elliott 2011). En sus más de ciento treinta páginas, editadas por el prestigioso sociólogo Michael Hout y por Stuart W. Elliott, se concluye que las pruebas estandarizadas no han acercado el nivel educativo de los Estados Unidos al de los países más exitosos. La mayor parte de estas pruebas, en el mejor de los casos, elevan mínimamente los resultados. Los exámenes externos para la obtención de la credencial de Secundaria Superior hacen disminuir el porcentaje de alumnos que consiguen tal título sin que por ello los niveles de aprendizaje mejoren.

En el caso norteamericano, la polémica se ha centrado en si el alumnado de aquellos estados que aplican pruebas externas obtienen mejores resultados que el de los que no las llevan a cabo. Los datos que arrojan diferentes estudios son contradictorios. Los más publicitados por los medios de comunicación muestran que los estados que han aplicado las reválidas puntúan peor en las evaluaciones de ámbito nacional que aquellos que no las han aplicado (Amrein y Berliner 2002). A ello hay que añadir que los resultados son aún peores para los estudiantes de menor estatus socioeconómico. En el caso de la reválida en Arizona, el noventa y siete por cien de los alumnos afroamericanos, latinos y nativos suspendieron en Matemáticas, porcentaje mucho mayor que el igualmente alto de los blancos.

Nichols (2007) explica el caso de Texas, cuyo ejemplo de reválidas fue en su momento considerado modélico y que, en realidad, no fue tal. Para comprobarlo simplemente bastó algo tan sencillo como comparar los resultados del examen de Texas (TAAS, *Texas Assessment of Academic Skills*) con el examen nacional NAEP (*National Assessment of Educational Progress*) para darse cuenta de que no había ningún milagro escolar en este estado. Los buenos resultados del TAAS eran consecuencia de que se estaba enseñando para preparar la prueba en las escuelas, así como que se impidió hacer el examen a los estudiantes menos académicos y de que se calculó erróneamente el número de personas que habían abandonado la escuela⁵. Jacob (2001) indicaba que las reválidas de Secundaria no tenían relación con los logros de los estudiantes. La única excepción era la de los alumnos con bajos resultados en lectura. Al igual que en otras investigaciones citadas, se detecta un mayor abandono escolar allí donde existen estos exámenes. Marchant y Paulson (2005) compararon los resultados de estas reválidas con el SAT y señalaron que los estados que las llevaban a cabo obtenían peores resultados en el SAT y además sus tasas de graduación en Secundaria eran más bajas.

Sin embargo, hay investigadores que discrepan radicalmente de estos resultados. Con los mismos datos que los utilizados por Amrein y Berliner (2002), Hanushek y Raymond (2004) llegaron a conclusiones diferentes, cuando no opuestas. Para ello, estos últimos incluyeron en su análisis estados inicialmente excluidos en el trabajo de los primeros. De este modo, descubrieron que los resultados en la NAEP eran más altos en los estados con *high-stakes* que en el resto. Del mismo modo que Hanushek y Raymond, Braun (2004) reanaliza los datos de Amrein y Berliner. Compara los resultados en cuarto y octavo cursos en la NAEP. Sus resultados ponen de manifiesto el impacto positivo de los *high-stakes* en los resultados escolares. Carnoy y Loeb (2002) analizan los resultados de la NAEP en el periodo compren-

dido entre 1996 y 2000 y detectan unos resultados moderadamente positivos en cuarto curso y muy favorables en octavo.

En su defensa, Amrein y Berliner aducen que no solo analizaron los datos de la NAEP, sino que también observaron los efectos que los exámenes de graduación en Secundaria Superior (*high school*) ejercían sobre los test de acceso a la universidad del tipo SAT o la participación en cursos de nivel avanzado. De hecho, los exámenes para la obtención de la Secundaria Superior han dado lugar a un incremento de las tasas de abandono escolar y a una mayor matriculación en el GED (*General Educational Development*, una certificación educativa para adultos)⁶, el cual es un sustituto de la credencial de Secundaria Superior.

Una revisión realizada en un texto de la OCDE (2012) concluye que las pruebas estandarizadas en los Estados Unidos parecen ser positivas para las matemáticas elementales y, sin embargo, no son estadísticamente significativas para la lectura ni dan lugar, ni de lejos, a las mejoras inicialmente previstas. Los resultados para los alumnos son inciertos.

En lo que sí hay pleno consenso en todos los estudios es en el dato de que los estudiantes de minorías étnicas salen peor parados que los de la mayoría blanca. Dianne Ravitch (2009; 2010) concluye que las diferencias entre los estudiantes de las minorías y los blancos apenas se han modificado. Pese a que los resultados de los alumnos de nueve y trece años han mejorado entre 2004 y 2008, el avance ha sido inferior al registrado en el periodo comprendido entre 1999 y 2004. Este tipo de test favorece a las familias acomodadas, ya que son las que se pueden pagar academias y tutores particulares para su preparación (Nichols y Berliner 2008b).

EQUIPOS DIRECTIVOS, PROFESORES Y ALUMNOS: ACTITUDES Y COMPORTAMIENTOS

¿Cuáles han sido las reacciones de los equipos directivos de los centros y de sus profesores y alumnos? En la medida en que estos test son relevantes para la propia existencia del centro escolar o para la promoción de su profesorado, se han producido numerosos casos de fraude consistente en evitar que se presentasen a las pruebas los estudiantes de menor rendimiento, cuya presencia haría bajar la nota media del centro o del estado. Thrupp y Hursh (2006) relatan el caso de Rodney Paige, superintendente del distrito escolar de Houston, quien ordenó a los directores bajo su jurisdicción que no consignaran como tales a los alumnos que habían abandonado los estudios. En su lugar, debían hacer constar que se habían trasladado a otro centro, que habían regresado a su país de origen o que adujeran alguna otra razón distinta del abandono escolar. Esta "contabilidad creativa" de Paige (quien entre 2001 y 2004 desempeñaría el cargo de secretario de Educación

en el gobierno Bush con el que sacó adelante la mencionada ley NCLB) redujo enormemente las estadísticas de fracaso escolar en Houston. Tal supuesto éxito le mereció distinciones públicas por la excelencia educativa conseguida bajo su mandato. Algo similar describía Mons (2009: 27) en el caso de Holanda.

"En 2006, en Holanda, un país con pruebas *high-stakes*, los inspectores investigaron prácticas de exclusión en ciertas escuelas como respuesta a los rumores relativos a los test realizados al final de la Educación Primaria. Se descubrió que, en algunos casos, los alumnos que tenían mayores probabilidades de ser enviados a los itinerarios escolares menos prestigiosos –*Leerwegondersteunend onderwijs* (grupos de refuerzo)– no hicieron el test. Una vez más, las acciones de los profesores llevaron a la obtención de unos resultados mejores de los que se deberían haber obtenido".

Los profesores terminan por etiquetar a su alumnado en función de los resultados previsibles en estas pruebas. En el trabajo cualitativo de Booher-Jennings (2005) se denunciaba que aquellos tendían a clasificar a los alumnos en tres grupos: los que aprobarían con seguridad, los susceptibles de mejora y los casos perdidos. La consecuencia era que los docentes concentraban su atención en los segundos ya que los resultados del resto son poco menos que inamovibles. Incluso, es frecuente que el profesorado ayude a los alumnos a hacer las pruebas, los cuales, a su vez, recurren a copiar y ayudarse entre sí. Todo ello desvirtúa considerablemente la validez de estos exámenes (Nichols y Berliner 2005). Amrein y Berliner (2003) citaban una encuesta que había revelado que el diez por ciento del profesorado admitía haber dado indicaciones a los alumnos para resolver las preguntas y el quince por ciento les concedió más tiempo del establecido. El libro de Crick, Jones y Hargrove (2003) recoge el dato de que el 90% del profesorado reconocía ser sometido a presión para que mejorase los resultados de sus alumnos.

Es obvio que estos embustes no justificarían la inexistencia de las pruebas externas. En todo caso, y debido a la casi imposibilidad de controlarlos, tales fraudes apuntan a la necesidad de conseguir altas dosis de consenso entre los diferentes estamentos que componen la comunidad educativa (carencia de acuerdo que ha sido especialmente criticada en el caso de la LOMCE).

Uno de los supuestos de las pruebas estandarizadas es que incrementarán la motivación (por lo menos la extrínseca) de los estudiantes. Aquí, una vez más, los resultados no son concluyentes. Algunas investigaciones detectan que los alumnos se sintieron más motivados, mientras que otras destacan que los profesores estaban preocupados por su menor amor por el aprendizaje (Mons 2007).

Un tema de gran relevancia es el del estrés que estas pruebas provocan en el alumnado. Sobre esta

cuestión, además de la consabida investigación en los Estados Unidos, contamos con datos del Reino Unido y de Australia (Polesel 2012). Cohen (2013) advertía que el 25% de los niños tiene trastornos de ansiedad, en buena medida derivados de las exigencias de la escuela. Honoré (2008) señalaba que el 41 % de los niños de entre nueve y trece años afirmó sentir estrés durante gran parte del tiempo o en todo momento porque tenían demasiadas tareas escolares.

CONCLUSIONES

Resulta difícil determinar qué cabría entender por un alumno educado. Posiblemente habría consenso en que queremos personas que lean periódicos, libros y otros materiales; que acudan a exposiciones artísticas de muy diverso tipo (desde el cine al teatro, pasando por museos, representaciones musicales y un largo etcétera); que, en la medida de sus posibilidades, viajen y conozcan otros mundos; que practiquen deporte; que sean capaces de comprender los rudimentos de las explicaciones científicas de nuestro mundo. Todo está muy bien, pero no basta. Es preciso también aprender a convivir, a amar al prójimo, a respetar a quienes no piensan como nosotros, a participar democráticamente en la vida de la polis. Y esto no se aprende escuchando a un profesor o leyendo libros: no queda más remedio que practicarlo cotidianamente en la escuela.

Junto a estas destrezas interpersonales, no habría que perder de vista, tal y como recordaba Levin (2006), habilidades intrapersonales como la perseverancia o el autocontrol. Nada de esto está en el punto de mira de las pruebas estandarizadas que han sido objeto de análisis en la literatura científica sobre el tema.

Son tantos, y tan sumamente graves, los problemas que plantean las pruebas externas que algunos de los mejores sistemas educativos, tras contrastarlo con la realidad empírica, simplemente las eliminan o las complementan con otras evaluaciones. A modo de ejemplo, a Australia –el país del mundo con mayor porcentaje de estudiantes universitarios– lo que le preocupa es la competencia, no tanto con la cada vez más marginal Europa, sino con los “tigres” asiáticos. En el informe Grattan (2012), se recoge una información valiosísima sobre qué ha sucedido con los

high-stakes tests australianos (los cuales se realizaban al finalizar la Primaria y la Secundaria Inferior y determinaban el tipo de escuela al que irían los alumnos). Estos exámenes han desaparecido y con ellos el énfasis en la mera adquisición repetitiva de conocimientos. Ahora se hacen exámenes *low-stakes* (algunos incluso han dejado de existir debido a la loable intención de rebajar el nivel de estrés del alumnado) cuyos resultados no se entregan a los estudiantes: sirven para que cada escuela pueda mejorar.

No está del todo claro que estas pruebas supongan una mejora del aprendizaje de los estudiantes. Más bien podría estar sucediendo todo lo contrario, ya que obligan a que el currículo se centre, no solo en las materias que preferentemente sean evaluadas (la Lengua y las Matemáticas) dejando al margen las demás, sino que cambian radicalmente el estilo docente, el cual se concentra en el aprendizaje meramente fáctico al margen de la experimentación, el contraste de ideas o la búsqueda de fuentes alternativas de información. En definitiva, y pese a que en un primer momento los niños de las clases media y alta salgan beneficiados, quedarán claramente incapacitados para desenvolverse en una sociedad en la que los conocimientos se multiplican cada poco tiempo exponencialmente (lo que obliga a aprender a lo largo de toda la vida) y en la que no sabemos cuáles serán los principales empleos que en el futuro habrá de desempeñar el alumnado que hoy está en la escuela.

La investigación educativa muestra un aumento de las desigualdades sociales y un incremento del abandono escolar al hilo de las pruebas externas, con lo que una vez más se estaría actuando contra una sociedad del conocimiento que exige elevar de un modo sostenido el nivel educativo del conjunto de la población.

Detrás de los *high stakes test*, al menos de los que se han venido utilizando habitualmente, parece haber un modelo educativo basado en la pasividad del alumnado, al que se le condena a aprender de un modo poco compatible con las exigencias del trabajo en equipo, de saber resolver imprevistos, de expresarse en público, del fomento de la creatividad y otras demandas cada vez más frecuentes en los empleos de la sociedad del conocimiento.

NOTAS

1. Literalmente significa “ningún niño puede quedar retrasado”. Se trata de una ley federal aprobada en 2001 que apoya la educación basada en estándares que son controlados por medio de exámenes externos propuestos por cada Estado. El interés de esta ley para el tema tratado en este texto es que supuso la consagración de los *high-stakes*, aunque, en este caso, las consecuencias recaen sobre los centros y sus profesores y no sobre el alumnado. La información oficial se puede ver en <http://www2.ed.gov/nclb/landing.jhtml>.
2. Véase el debate parlamentario celebrado en el pleno del Congreso de los Diputados el 16 de Julio de 2013 (http://www.congreso.es/public_oficiales/L10/CONG/DS/PL/DSCD-10-PL-129.PDF).
3. Se puede ver en el siguiente link del Ministerio de Educación: <http://www.mecd.gob.es/servicios-al-ciudadano->

- mecd/dms/mecd/servicios-al-ciudadano-mecd/informacion-publica/audiencia-informacion-publica/cerrados/2015/evaluaciones-finales/proyecto-RD-evaluaciones-finales/proyecto%20RD%20evaluaciones%20finales.pdf
4. TIMMS es el acrónimo de *Trends in International Mathematics and Science Study* (Estudio de las Tendencias en Matemáticas y Ciencias). Es una evaluación internacional de conocimientos de estas dos materias de los estudiantes de los grados cuarto y octavo en diferentes países.
5. De hecho en este caso se habla del efecto *Lake Wobegon* (nombre de una ciudad ficticia de los Estados Unidos) y que consiste en sobreestimar las capacidades propias.
6. Se puede ver una explicación en <http://www.cde.ca.gov/ta/tg/gd/>

REFERENCIAS BIBLIOGRÁFICAS

- Abell Foundation. 2003. High Risk or High Time? A Critical Junction in implementing Maryland's High-Stakes High School Assessment as a Graduation Requirement. History and background; problems and recommendations. Informe julio 2003. Baltimore: The Abell Foundation. Consulta 2 de junio de 2013 (http://www.abell.org/sites/default/files/publications/ed_highstakes_703.pdf).
- Amrein, A. L. y D. C. Berliner. 2002. "High-stakes testing, uncertainty, and student learning". *Education Policy Analysis Archives* 10(18). Consulta 4 de agosto de 2014 (<http://epaa.asu.edu/ojs/article/viewFile/297/423>).
- Amrein, A. L. y D. C. Berliner. 2002. *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*. Tempe: Arizona State University. Consulta 3 de abril de 2012 (<http://nepc.colorado.edu/files/EPSSL-0211-125-EPRU.pdf>).
- Amrein, A. L. y D. C. Berliner. 2003. "The effects of High-stakes testing on student motivation and learning". *Educational Leadership*, Febrero 2003: 32-38. Consulta 22 de marzo de 2014 (http://www.wou.edu/~girodm/611/testing_and_motivation.pdf).
- Anagnostopoulos, D. 2003. "Testing and student engagement with literature in urban classrooms: a multi-layered perspective". *Research in the Teaching of English* 38: 177-212.
- Apple, M. W. y J. A. Beane. 2005. *Escuelas democráticas*. Madrid: Morata.
- Au, W. 2008. "Devising inequality: a Bernsteinian analysis of high-stakes testing and social reproduction in education". *British Journal of Sociology of Education* 29: 639-651. <http://dx.doi.org/10.1080/01425690802423312>
- Au, W. 2009. *Unequal By Design. High-Stakes Testing and the Standardization of Inequality*. Nueva York: Routledge.
- Bishop, J. B. 2005. "High School Exit Examinations: When Do Learning Effects Generalize?" CAHRS Working Paper Series. School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, Cornell University, Ithaca. Consulta 6 de junio de 2015 (<http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1003&context=cahrswp>).
- Bishop, J. B. 2006. "Drinking from the fountain of knowledge: Student incentive to study and learn-Externalities, information problems, and peer pressure". Pp. 909-944 en *Handbook of the Economics of Education*, editado por E. A. Hanushek, S. J. Machin y L. Woessmann. Amsterdam: North-Holland.
- Booher-Jennings, J. 2005. "Below the bubble: 'Educational triage' and the Texas accountability System", *American Educational Research Journal* 42: 231-268. <http://dx.doi.org/10.3102/00028312042002231>
- Braun, H. 2004. "Reconsidering the Impact of High-stakes Testing". *Education Policy Analysis Archives* 12(1). <http://dx.doi.org/10.14507/epaa.v12n1.2004>
- Campbell, D. T. 1976. *Assessing the Impact of Planned Social Change*. Hanover (New Hampshire): The Public Affairs Center, Dartmouth College.
- Carnoy, M. y S. Loeb. 2002. "Does External Accountability Affect Student Outcomes?: A Cross-State Analysis". *Educational Evaluation and Policy Analysis* 24: 305-331. <http://dx.doi.org/10.3102/01623737024004305>
- Cohen, L. J. 2013. "The Drama of the Anxious Child". *Time* (26 de septiembre de 2013). Consulta 30 de julio de 2014 (<http://ideas.time.com/2013/09/26/the-drama-of-the-anxious-child/#ixzz2g0AHZCBF>).
- Crocker, L. 2005. "Teaching For the Test: How and Why Test Preparation Is Appropriate". Pp. 159-174 en *Defending Standardized Testing*, editado por R. P. Phelps. New Jersey: Lawrence Erlbaum Associates Publishers.
- Gail Jones, M; B. D. Jones y T. Y. Hargrove. 2003. *The Unintended Consequences of High-Stakes Testing*, Lanham (Maryland): Rowman & Littlefield.
- Gardner, H. 2002. *La educación de la mente y el conocimiento de las disciplinas. Lo que todos los estudiantes deberían comprender*. Barcelona: Paidós.
- Garicano, L. 2015. "¿El fin de la clase magistral?". *El País*, 2 de junio de 2015. Consulta el 2 de junio de 2015 (http://economia.elpais.com/economia/2015/03/05/actualidad/1425575098_421184.html).
- Jensen, B., A. Hunter, J. Sonnemann and T. Burns. 2012. *Catching up: Learning from the best schools systems in East Asia*. Melbourne: Grattan Institute. Consulta 12 de abril de 2013 (http://grattan.edu.au/wp-content/uploads/2014/04/129_report_learning_from_the_best_main.pdf).
- Hanushek, E. A. y M. E. Raymond. 2004. "Does School Accountability Lead to Improved Student Performance?". *Journal of Policy Analysis and Management* 24: 297-327. <http://dx.doi.org/10.3386/w10591>
- Honoré, C. 2008. *Bajo presión. Cómo educar a nuestros hijos en un mundo hiperexigente*. Barcelona: RBA.
- Hout, M. y S. W. Elliott. 2011. "Incentives and Test-Based Accountability in Education Committee". Washington, D.C.: National Academies Press. Consulta 30 de Julio de 2014 (http://www.nap.edu/catalog.php?record_id=12521).
- Huang, Min-Hsiung. 2009. "Do Public Exams Raise Student Performance? A Cross-National Difference-in-Differences Analysis". *Journal of Social Sciences and Philosophy* 21: 1-24.
- Jacob, B. A. 2001. "Getting tough? The impact of high school graduation exams". *Educational Evaluation and Policy Analysis* 23: 99-121. <http://dx.doi.org/10.3102/01623737023002099>
- Klein, D. 2010. *The Differing Purposes and Effects of State-Wide Exit Examinations A comparison of Finland, Ireland and the Netherlands from a Governance Perspec-*

- five. Trabajo presentado en la 36ª Reunión Anual de la Asociación Internacional de Evaluación de la Educación (OIEA), julio, Bangkok, Tailandia. Consulta 30 de mayo de 2014 (http://www.iaea.info/documents/paper_4d519c43.pdf).
- Marchant, G. J. y S. E. Paulson. 2005. "The relationship of high school graduation exams to graduation rates and SAT scores". *Education Policy Analysis Archives* 13: 2-17. Consulta 30 de julio de 2014 (<http://epaa.asu.edu/ojs/article/view/111/237>).
- Mons, N. 2007. *Les nouvelles politiques éducatives*. París: PUF. <http://dx.doi.org/10.3917/puf.mons.2007.01>
- Mons, N. 2009. "Theoretical and real effects of standardised assessment". Background paper to the study *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results* EACEA; Eurydice. Consulta 30 de julio de 2014 (http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111EN.pdf).
- Morris, A. 2011. "Student Standardised Testing: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers 65, Organisation for Economic Co-operation and Development, Paris. Consulta 30 de julio de 2014 (<http://dx.doi.org/10.1787/5kg3rp9qbnr6-en>).
- Nichols, S. L. 2007. "High-Stakes Testing: Does It Increase Achievement?". *Journal of Applied School Psychology* 23: 109-128. http://dx.doi.org/10.1300/J370v23n02_04
- Nichols, S.L. y D. C. Berliner 2005. *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing*. Tempe: Arizona State University. Consulta 30 de julio de 2014 (<http://files.eric.ed.gov/fulltext/ED508483.pdf>).
- Nichols, S. L. y D. C. Berliner. 2008a. *Collateral Damage. How high-stakes testing corrupts America's school*. Cambridge: Harvard Education Press.
- Nichols, S. L. y D. C. Berliner. 2008b. "Why Has High-Stakes Testing So Easily Slipped into Contemporary American Life?". *Phi Delta Kappa* 89: 672-676. Consulta 2 de junio de 2013 (<http://www.jstor.org/stable/20442603>).
- OECD. 2011. *Equity and Quality in Education - Supporting Disadvantaged Students and Schools*. Consulta 2 de junio de 2013 (<http://www.oecd.org/edu/school/equityandqualityineducation-supportingdisadvantagedstudentsandschools.htm>).
- Plake, B. S. 2005. "Doesn't Everybody Know That 70% is Passing?". Pp. 175-186 en *Defending Standardized Testing*, editado por R. P. Phelps. New Jersey: Lawrence Erlbaum Associates Publisher.
- Polesel, J., N. Dulfer y M. Turnbull. 2012. "The Experience of Education: The impacts of high stakes testing on school students and their families". *Literature Review* January 2012. Whitlam Institute. Consulta 30 de julio de 2014 (http://www.whitlam.org/_data/assets/pdf_file/0008/276191/High_Stakes_Testing_Literature_Review.pdf).
- Ravitch, D. 2009. "Time to Kill 'No Child Left Behind'". *Education Week* 28(33). Consulta 2 de julio de 2013 (http://www.edweek.org/ew/articles/2009/06/04/33ravitch_ep.h28.html).
- Ravitch, D. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Nueva York: Basic Books.
- Reich, R. B. 2001. "Standards for what?". *Education Week* 20: 64-65.
- Roderick, M. y J. Nagaoka 2005. "Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless?". *Educational Evaluation and Policy Analysis* 27: 309-340. <http://dx.doi.org/10.3102/01623737027004309>
- Ruiz i Tarragó, F. 2010. "Los centros educativos ante la rendición de cuentas: 'lessons from America'". *La evaluación en los procesos de mejora y autonomía* 5: 14-17. Consulta 30 de julio de 2014 (<http://www.xtec.cat/~fruiz/present/D1479.pdf>).
- Sireci, S. 2005. "The Most Frequently Unasked Questions About Testing". Pp. 11-122 en *Defending Standardized Testing*, editado por R. P. Phelps. New Jersey: Lawrence Erlbaum Associates Publishers.
- Thrupp, M. y D. Hursh 2006. "The Limits of Managerialist School Reform: The Case of target-Setting in England and the USA". Pp. 642-653 en *Education, Globalization & Social Change*, editado por H. Lauder, P. Brown, J. A. Dillabough and A. H. Halsey. Oxford: Oxford University Press.
- Wagner, T. 2010. *The Global Achievement Gap. Why even our best schools don't teach the new survival skills our children need and what we can do about it?* Nueva York: Basic Books.

RAFAEL FEITO ALONSO es catedrático de Sociología en la Facultad de Ciencias Políticas de la Universidad Complutense de Madrid. Es autor de *Nacidos para perder* (Madrid, CIDE. 1990), *Estructura social contemporánea* (Madrid, Siglo XXI, 1995), *Clases sociales y comportamiento político en España* (Madrid, Entinema, 1998), *Los retos de la educación obligatoria* (Barcelona, Ariel, 2000), *Una educación de calidad para todos* (Madrid, Siglo XXI, 2002), *Otra escuela es posible* (Madrid, Siglo XXI, 2006), *Construyendo escuelas democráticas* (Hipatia, Barcelona, 2008), *Los retos de la participación escolar* (Madrid, Morata, 2011) y de más de un centenar de artículos en revistas especializadas y generalistas.