

# 11

## **OPINIÓN DEL PROFESORADO SOBRE LAS FORTALEZAS Y DEBILIDADES DE LAS EVALUACIONES EXTERNAS DE APRENDIZAJE EN ESPAÑA<sup>1</sup>**

**(THE OPINION OF TEACHERS ON THE STRENGTHS AND WEAKNESSES OF HIGH-STAKES TESTS IN SPAIN)**

Héctor Monarca  
María Fernández-Agüero  
*Universidad Autónoma de Madrid*

*DOI: 10.5944/educXX1.19416*

### **Cómo referenciar este artículo/How to reference this article:**

Monarca, H. y Fernández-Agüero, M. (2018). Opinión del profesorado sobre las fortalezas y debilidades de las evaluaciones externas de aprendizaje en España. *Educación XX1*, 21(2), 249-273, doi: 10.5944/educXX1.19416

Monarca, H. & Fernández-Agüero, M. (2018). Opinión del profesorado sobre las fortalezas y debilidades de las evaluaciones externas de aprendizaje en España. [The opinion of teachers on the strengths and weaknesses of high-stakes tests in Spain]. *Educación XX1*, 21(2), 249-273, doi: 10.5944/educXX1.19416

## **RESUMEN**

El objetivo de este artículo es presentar las fortalezas y debilidades que, según la opinión del profesorado de centros de educación primaria y educación secundaria obligatoria, tienen las evaluaciones externas de España. Para ello se han analizado las dos preguntas abiertas de un cuestionario más amplio orientadas a identificar las opiniones sobre estas fortalezas y debilidades. Del total de respuestas obtenidas se han seleccionado las 1.607 correspondientes a 5 Comunidades Autónomas repartidas de la siguiente manera: 321 de Andalucía, 445 de Cataluña, 239 de Galicia, 491 de Madrid y 111 del País Vasco. La aplicación del cuestionario se inicia poco antes de aprobarse la nueva Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE), aunque

el debate en torno a ella ya llevaba al menos un año dentro de la comunidad educativa, un rasgo que define el contexto en el que se desarrolló la encuesta. Las respuestas obtenidas fueron codificadas y categorizadas mediante un proceso inductivo-deductivo que ha seguido la siguiente secuencia: I- Fuente de Información, II- Códigos emergentes, III- Categorías y IV- Dimensiones. La información ha sido tratada de acuerdo con las siguientes variables: a) «Comunidad Autónoma», b) «Etapa Educativa» (Educación Primaria y Educación Secundaria Obligatoria), y c) «Función principal desempeñada» en el centro por quien ha respondido la encuesta (miembro del equipo directivo y resto de profesores). Los resultados muestran que, en términos generales, hay mayor porcentaje de respuestas que hacen mención a efectos negativos que a positivos. Dentro de las fortalezas destaca la función diagnóstica que pueden tener las evaluaciones externas y dentro de las debilidades destaca especialmente la descontextualización de estas pruebas. El análisis muestra importantes diferencias por comunidad autónoma y, en algunos casos, también por función principal desempeñada por el que ha contestado la encuesta.

## **PALABRAS CLAVE**

Política educativa; pruebas educativas; evaluaciones externas; rendición de cuentas.

## **ABSTRACT**

The aim of this paper is to identify the strengths and weaknesses of high-stakes examinations in the view of Primary and Secondary school teachers in Spain. For that purpose, we have analyzed the answers to two open questions from a larger survey designed to determine the teachers' opinions on these strengths and weaknesses. Out of the total sample, we have selected 1,607 responses corresponding to five Autonomous Communities or regions, distributed as follows: 321 from Andalusia, 445 from Catalonia, 239 from Galicia, 491 from Madrid and 111 from the Basque Country. The questionnaire was applied between November 2013 and May 2014. Its administration began shortly before the new Organic Law for the Improvement of Quality in Education (LOMCE) was passed, although the debate around it had already been taking place for at least a year in the educative community, an important trait of the context in which the survey was carried out. Responses were coded and categorized applying an inductive-deductive process with this sequence: I. Source of information; II. Emerging codes; III. Categories; and IV. Dimensions. The information has been treated according to the following variables: (a) Autonomous Communities; (b) education stage (Primary Education and Compulsory Secondary Education); and (c) main role undertaken by respondents at the school (members of the school managing team vs. the rest of the teachers). Results show that by and large there are more references to weaknesses than to strengths. Also, the diagnostic function of tests and their decontextualisation stand out as the most frequent strength and weakness

respectively. The analysis reveals significant differences between autonomous communities and also, in some cases, differences depending on the role undertaken by respondents.

## **KEY WORD**

Educational policy; educational testing; high stakes tests; accountability.

## **INTRODUCCIÓN**

Las evaluaciones externas de aprendizaje son cada vez más usadas como herramientas o dispositivos diseñados fuera del ámbito escolar con diversos propósitos (Eurydice, 2009). Por lo general con la intención explícita de comprobar determinados logros o competencias de los estudiantes a ciertas edades o momentos de su escolarización, ya sea por escuelas, países o regiones.

El aumento de estas evaluaciones ha ido acompañado por la proliferación de investigaciones y producción académica sobre la temática. La misma refleja una gran variedad de abordajes y perspectivas teóricas: como instrumentos de políticas (Conley, 2015; Feniger, Israeli y Yehuda, 2015; Skedsmo, 2011), como instrumentos para la mejora escolar (Domenech, Blazquez, Poza y Muñoz-Miquel, 2015; Hopfenbeck, Flórez y Tolo, 2015; Turnipseed y Darling-Hammond, 2015); como nuevas formas de gobernanza (Fernández-González y Monarca, 2018; Parcerisa y Falabella, 2017; Rezai-Rashti, Segeren y Martino, 2016); como dispositivos para la configuración de nuevos significados sobre la enseñanza, el aprendizaje o la educación en general (Monarca, 2015; Pettersson, Popkewitz y Lindblad, 2016; Simmonds y Webb, 2013), etc.

Más allá de simplificaciones analíticas, la realidad nos muestra una gran complejidad difícil de clasificar. De esta manera, mientras que ciertas perspectivas cuestionan profundamente la racionalidad inherente a estos dispositivos (Komatsu y Rappleye, 2017; Pettersson, Popkewitz y Lindblad, 2016), otras los defienden y promueven, entendiendo que aportan o podrían aportar beneficios para el sistema educativo y que son instrumentos potenciales para la mejora de las escuelas, la enseñanza y el aprendizaje (Darling-Hammond y Falk, 2013; Flórez, 2015; Hopfenbeck et al., 2015; Smith, 2014).

En este trabajo se presentan las fortalezas y debilidades de las evaluaciones externas según la opinión del profesorado de centros de educación primaria y educación secundaria obligatoria de Andalucía,

Cataluña, Galicia, Madrid y País Vasco. Para ello se analizan las dos preguntas abiertas sobre este aspecto en las 1.607 encuestas respondidas en estas Comunidades Autónomas (CCAA). Estas evaluaciones fueron introducidas de forma generalizada en España en el año 2006 por la Ley Orgánica de Educación (LOE) y han sido aumentadas por la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE) en el año 2013.

## **FORTALEZAS DE LAS EVALUACIONES EXTERNAS**

En términos generales una serie de autores considera que existe consenso en que las evaluaciones externas de aprendizaje sostenidas a lo largo del tiempo contribuyen a mejorar la enseñanza y el aprendizaje (Darling-Hammond y Falk, 2013; Eurydice, 2009; Flórez, 2015; Hopfenbeck et al., 2015; Smith, 2014). Desde esta perspectiva se afirma que contribuyen a: a) promover y aumentar el aprendizaje, fundamentalmente cuando existe una rápida retroalimentación (Domenech et al., 2015); b) diagnosticar y seguir los procesos de enseñanza-aprendizaje (Skedsmo, 2011); c) consolidar y dirigir el aprendizaje, en tanto contribuyen a que los docentes y los estudiantes dediquen más tiempo a ciertos contenidos (Harris y Herrington, 2006); y d) incentivar el estudio, en tanto motivan a los alumnos y alumnas (Bishop, 2006). Por su parte, Stobart y Eggen (2012) sostienen que algunos autores ven positivamente que estas evaluaciones hacen que los docentes trabajen más duro y con más eficacia para cubrir más materia.

Otros afirman que las pruebas nacionales en sí mismas pueden ser un referente para los centros y que el uso comparativo de sus resultados puede ser usado para la mejora de los mismos (Mons, 2009). En este sentido, Thélot (2002) afirma que en ocasiones las pruebas nacionales son usadas como un espejo en el cual el profesorado puede mirar el resultado de sus acciones. Desde una lógica similar, Darling-Hammond y Falk (2013) consideran que la participación del profesorado en los procesos de corrección, aplicación y muy especialmente de reflexión sobre los resultados de estas evaluaciones, aportan importantes beneficios para la mejora de la enseñanza y del aprendizaje.

Por otra parte, autores como Conley (2015) o Turnipseed y Darling-Hammond (2015), destacan los potenciales beneficios de lo que denominan evaluaciones «profundas o auténticas», orientadas hacia competencias complejas, de orden superior. Aunque simultáneamente reconocen que no son evaluaciones que abundan.

Finalmente, algunos autores valoran los datos que las evaluaciones externas aportan en tanto consideran que favorecen o permiten actuaciones

basadas en evidencias (Coburn, Hill y Spillane, 2016; Kowalski y Lasley, 2009; Spillane, 2012). En este sentido, los datos que las pruebas externas ofrecen serían útiles para conocer y comprender el funcionamiento de las escuelas, la enseñanza y el aprendizaje y así poder tomar las decisiones adecuadas para la mejora de los mismos.

Es preciso destacar que en algunos casos estos supuestos beneficios son matizados por los mismos autores. Muchos de ellos cuando se refieren a las fortalezas o beneficios de este tipo de evaluaciones lo hacen en relación a una serie de condiciones o de características que deberían tener, pero que reconocen que en muchos casos no se cumplen.

## **DEBILIDADES DE LAS EVALUACIONES EXTERNAS**

En una revisión de la literatura sobre las debilidades de las evaluaciones externas es posible mencionar la reducción o empobrecimiento del currículo o la enseñanza (Feniger et al., 2015; Klenowski y Wyatt-Smith, 2012; Monarca, 2012; Polesel, Rice y Dulfer, 2014; Winter, 2017). En esta línea se cuestionan las competencias o conocimientos que los exámenes ponen en valor o, por el contrario, quedan devaluados o excluidos (Conley, 2015; Geyer, 2012; Hopfenbeck et al. 2015; Smith, 2014; Turnipseed y Darling-Hammond, 2015).

Con respecto específicamente a la forma de enseñar, muy condicionada por los tipos de pruebas externas que tendrán que realizar los estudiantes, se ha destacado la manera fragmentada y poco integrada en que los contenidos son enseñados (Conley, 2015; Monarca, 2012;). Junto con esto, se menciona el tipo de actividades que se realizan y el tipo de competencias cognitivas simples que se estimulan (Klenowski y Wyatt-Smith, 2012), muy alejadas del pensamiento de orden superior. De esta manera, se produciría un proceso de empobrecimiento de las prácticas docentes, de estandarización de las mismas (Geyer, 2012; Polesel, Dulfer y Turnbull, 2012). En esta misma línea cabe destacar la amplia evidencia existente a nivel internacional sobre la enseñanza centrada en las pruebas (Conley, 2015; Hopfenbeck et al. 2015; Klenowski y Wyatt-Smith, 2012; Turnipseed y Darling-Hammond, 2015).

Otros autores han señalado que las evaluaciones externas aumentan o pueden aumentar las desigualdades, destacando los efectos sobre las minorías étnicas o grupos sociales en desventaja (Falabella, 2014; Smith, 2014). Este aspecto ha sido relacionado con la concentración de los recursos en ciertas escuelas (Bourke y Mentis, 2013; Koch y DeLuca, 2012), materias o personas (Falabella, 2014; Feniger et al., 2015; Klenowski y Wyatt-Smith, 2012; Smith, 2014; Sun, Saultz y Ye, 2016). Por otra parte, se menciona que

influyen en la diferenciación, jerarquización, etiquetación y estigmatización de centros educativos y estudiantes (Bourke y Mentis, 2013; Monarca, 2012; Polesel et al., 2014; Simmonds y Webb, 2013), muy especialmente de aquellos que tienen necesidades educativas especiales (Bacon, 2015), o situaciones de desventaja social.

Con respecto a los docentes y su profesionalidad, la literatura ha señalado algunos efectos negativos de las evaluaciones externas que influyen en el malestar docente. En este sentido, se han relacionado las políticas de rendición de cuentas basadas en exámenes con sentimientos de contradicción, tensión y ansiedad en los docentes, que generan lo que se ha denominado conflicto de rol, a partir del cual aumentan los sentimientos de desprofesionalización (Polesel et al., 2012; Smith, 2014; Wilkins, 2015; Winter, 2017).

Se han señalado también debilidades relacionadas con la información que se genera a partir de estas evaluaciones, sobre cómo se comunica y sobre los usos que se hace de la misma. Al respecto, Koch y DeLuca (2012) sostienen que los resultados pueden ser interpretados de diversas maneras según el clima político o los intereses de los diversos actores. En este sentido, Skedsmo (2011) cuestiona el uso indiscriminado de la información que se genera a partir de estas pruebas, sin atender a las necesidades específicas de información de los distintos actores: estudiantes, familias, profesorado, directivos y políticos. Por otra parte, Conley (2015), al igual que Klenowski y Wyatt-Smith (2012), considera que la información que se genera a partir de los exámenes es limitada o poco útil para orientar la práctica de los docentes, en tanto suele contener escasa o ninguna recomendación para hacer frente a los problemas que se detectan. Por último, una gran variedad de autores han cuestionado la publicación de los resultados y la elaboración de rankings a partir de los mismos, aludiendo a los efectos que ello genera (Feniger et al., 2015; Koch y DeLuca, 2012; Monarca, 2012, 2015; Polesel et al., 2012; Simmonds y Webb, 2013).

Finalmente es preciso destacar aquellos cuestionamientos que se realizan a la racionalidad subyacente a las evaluaciones externas (Feniger et al. 2015; Komatsu y Rappleye, 2017; Pettersson, Popkewitz y Lindblad, 2016). Según esta perspectiva, una racionalidad performativa basada en los números para la gestión y gobierno de prácticas complejas que contribuye a la mercantilización de la educación (Sun et al., 2016).

## **METODOLOGÍA**

El presente trabajo se desprende de una investigación más amplia sobre las evaluaciones externas en España realizada a partir de un enfoque

metodológico mixto cuanti-cualitativo. En este caso se centra en el siguiente interrogante: ¿cuáles son las fortalezas y debilidades de las evaluaciones externas desde el punto del profesorado de centros de Educación Primaria y Educación Secundaria Obligatoria de España?, el cual fue abordado a través de dos preguntas abiertas realizadas en el marco de una encuesta.

El cuestionario fue enviado mediante correo electrónico entre noviembre de 2013 y mayo de 2014 al universo completo de centros de educación primaria y de educación secundaria obligatoria de España, tanto públicos como concertados-privados. Se han seleccionado para el análisis las 1.607 respuestas abiertas obtenidas en 5 CCAA, repartidas de la siguiente manera: 321 de Andalucía, 445 de Cataluña, 239 de Galicia, 491 de Madrid y 111 del País Vasco. Estas CCAA han sido escogidas por una combinación de criterios: a) existencia de otras evaluaciones externas propias además de las establecidas en la normativa estatal, b) su importancia histórica y económica, y c) existencia de otra lengua cooficial.

En el análisis de la información se han tenido en cuenta las siguientes variables: a) «Comunidad Autónoma», b) «Etapa Educativa» (Educación Primaria y Educación Secundaria Obligatoria), y c) «Función principal desempeñada» (miembro del equipo directivo y resto de profesores) (ver tabla 1).

Tabla 1  
*Características de la muestra*

CCAA	N.º encuestas						Totales
	Educación Primaria			Educación Secundaria Obligatoria			
	ED	Prof	Subtotales	ED	Prof	Subtotales	
Andalucía	125	45	170	73	78	151	321
Cataluña	141	53	194	116	135	251	445
Galicia	52	58	110	40	89	129	239
Madrid	101	119	220	74	197	271	491
País Vasco	36	22	58	33	20	53	111
Totales	455	297	752	336	519	855	1.607

Nota: elaboración propia. Leyenda: ED=miembros del equipo directivo; Prof=profesorado de cualquier materia del currículo y de apoyo a necesidades educativas especiales.

## PROCEDIMIENTO DE CODIFICACIÓN-CATEGORIZACIÓN DE LAS RESPUESTAS ABIERTAS

Siguiendo a Basit (2003), las respuestas obtenidas fueran tratadas de forma inductiva-deductiva de acuerdo con la siguiente secuencia: I- Fuente de Información, palabras o frases breves ofrecidas como respuestas por los informantes; II- Códigos emergentes, primer sistema de significados que surgen del texto-respuesta; III- Categorías, proceso de agrupamiento por abstracción que supone reducir los códigos emergentes a unidades de significados de mayor potencialidad explicativa (Flick, 2006), IV- Dimensiones, nuevo proceso de agrupamiento por abstracción, en este último caso, claramente orientado por el marco teórico existente sobre la temática (Flick, 2006).

- *Primera fase.* A partir de la sugerencia de Cohen y Manion (1989), dos investigadores realizaron una primera lectura de una parte de las respuestas obtenidas, entre el 20 y el 25% de las mismas en cada comunidad. A partir de ello, cada investigador realizó una primera lista de códigos según el significado que emergía de las respuestas dadas por los encuestados. De esta manera, cada vez que aparecía un nuevo significado se generaba un código.
- *Segunda fase.* Elaborada esta primera lista de códigos por cada codificador de forma independiente, los mismos fueron analizados conjuntamente por ambos, atendiendo a las reglas asociadas a la construcción de códigos: homogeneidad, exhaustividad, exclusión mutua, claridad, objetividad, utilidad y pertinencia (Espín, 2002). Se prestó especial atención a la «consistencia», es decir a la aplicación del «mismo criterio a lo largo de todo el análisis» por parte de ambos codificadores (Espín, 2002, p. 103). En esta fase fue necesario realizar algunos ajustes y acuerdos de aplicación de criterios para generar una única lista de códigos.
- *Tercera fase.* Una vez realizada la tarea anterior, se procedió a elaborar categorías mediante un proceso de agrupamiento por abstracción, prestando especial atención a su validez, es decir, a su relación directa con el objetivo de análisis del estudio (Basit, 2003): la identificación de las fortalezas y debilidades de las evaluaciones externas de aprendizaje. En este caso se tuvo en cuenta el marco teórico para la formulación o reformulación de los códigos generados y ajustados en la fase 2.
- *Cuarta fase.* Las categorías fueron agrupadas en dimensiones, al igual que en el caso anterior, mediante un proceso de agrupamiento de mayor abstracción orientado exclusivamente por el marco teórico.



Realizado esto, se procedió a la cuantificación de las categorías (Hsieh y Shannon, 2005), atendiendo exclusivamente a la frecuencia de aparición (Espín, 2002). La cuantificación ha tenido como objetivo mostrar el «peso» que, dentro de los «relatos» —las respuestas a las dos preguntas abiertas—, tienen las distintas opiniones, ya codificadas y reflejadas en las categorías mencionadas en la fase 2. Por tanto, tal como se hace en el trabajo de Brom, Dalle y Elbert (2007), los resultados se presentan combinando el peso de las opiniones de los encuestados en términos de porcentajes, con fragmentos de sus relatos que enriquecen la descripción y comprensión de las opiniones ofrecidas.

## **RESULTADOS**

Los resultados se presentan de forma diferenciada, por un lado las fortalezas y por otro las debilidades, reflejándolos tanto de forma global como por CCAA y mostrando también las particularidades según las variables contempladas. Tal como se puede apreciar en la Figura 1, del proceso de codificación de las fortalezas se han generado 10 categorías finales que se han agrupado en 3 dimensiones. Por otra parte, en el proceso de codificación de las debilidades se han generado 15 categorías finales que se han agrupado en 4 dimensiones.

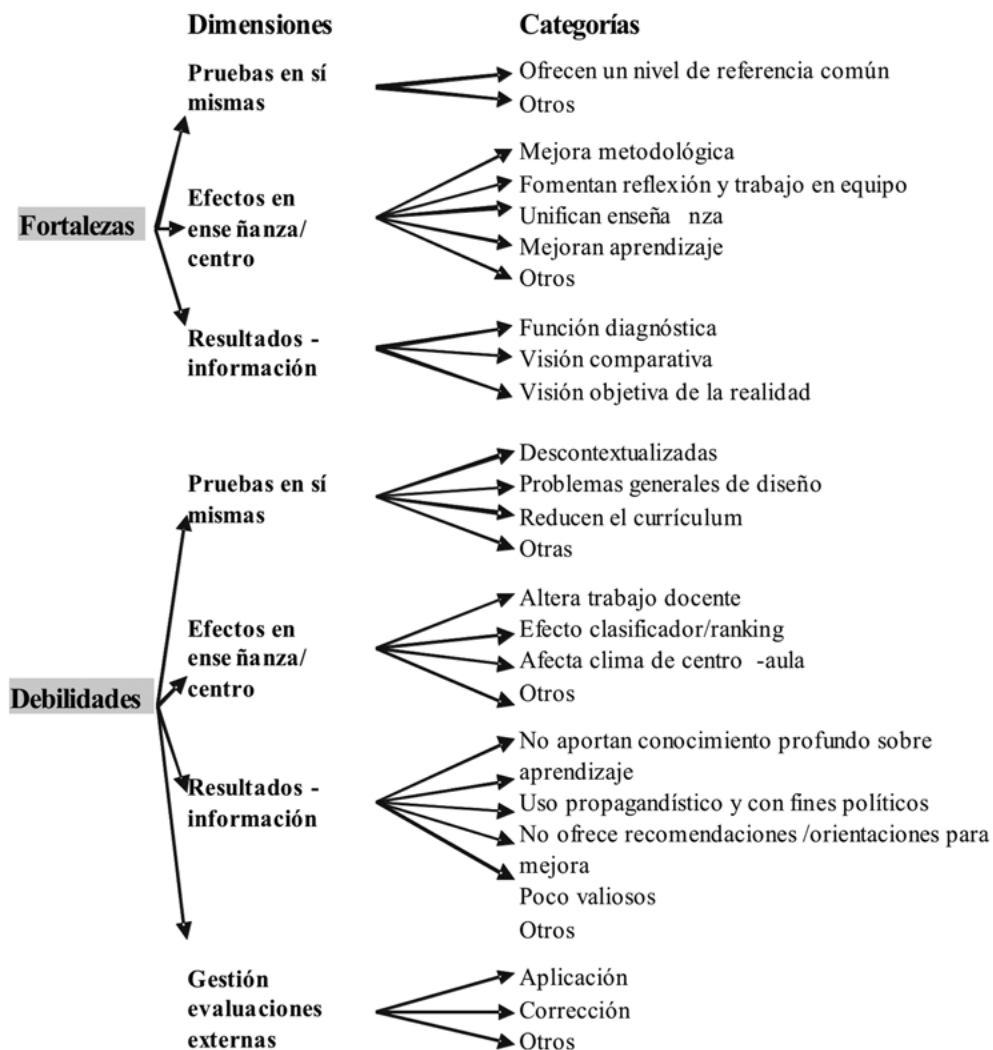


Figura 1. Resultados del proceso de codificación

## FORTALEZAS DE LAS EVALUACIONES EXTERNAS

Antes del detalle de los resultados es preciso mencionar que el 33% del total de la muestra ha contestado que las evaluaciones externas no tienen ninguna fortaleza: Andalucía 32%, Cataluña 45%, Galicia 13%, Madrid 49% y País Vasco 18%; con diferencias según la función principal desempeñada: no las hay para el 44% de los profesores y el 21% de los miembros de equipo directivo. Apenas existe diferencia por etapa educativa. A continuación se

presenta en la tabla 2 un resumen general de las opiniones sobre fortalezas de las evaluaciones externas de aprendizaje por CCAA y muestra total; posteriormente se ofrecen los resultados por dimensión, categoría y variables contempladas.

Tabla 2  
*Opinión sobre las fortalezas de las evaluaciones externas (porcentajes)*

Dimensiones	Categorías	Por CCAA					Total Muestra
		A	C	G	M	PV	
		N321	N445	N239	N491	N111	N1607
Pruebas en sí mismas	Ofrecen un nivel de referencia común	11	15	10	6	11	10
	Otros	1	2	1	3	0	2
	Total dimensión	12	17	11	9	11	12
Efectos en enseñanza-centro	Mejora metodológica	18	19	12	8	24	15
	Fomentan reflexión y trabajo en equipo	5	9	6	2	14	6
	Unifican enseñanza	3	8	3	6	2	5
	Mejoran aprendizaje	6	4	3	5	6	5
	Otros	0	2	2	1	0	1
Total dimensión	32	42	26	22	46	32	
Resultados-información	Función diagnóstica	26	27	14	15	27	21
	Visión comparativa	16	24	5	9	31	16
	Visión objetiva de la realidad	14	14	10	9	21	12
	Total dimensión	56	65	29	33	79	49

Nota: elaboración propia. Los datos indican qué porcentaje del total de cada CCAA o muestra total ha mencionado esa categoría y dimensión. Leyenda: A=Andalucía; C=Cataluña; G=Galicia; M=Madrid; PV=País Vasco.

### **Relacionadas con «resultados/información»**

Las principales fortalezas que se han destacado en el total de la muestra se refieren a la dimensión «resultados/información», señalada por el 49% de los encuestados. En términos generales, las opiniones de esta dimensión aparecen con más frecuencia en los encuestados del País Vasco (79%), seguidos por los de Cataluña (65%) y por los de Andalucía (56%), y en menor medida por los de Madrid (33%) y Galicia (29%). Una tendencia similar se aprecia en cada una de las categorías que componen esta dimensión.

De estos, la mayoría se refiere específicamente a su «función diagnóstica» (21%), en cuanto al conocimiento que aportan del alumnado, la enseñanza y el centro. Tal como afirman los encuestados «ayudan a detectar aquellas áreas que se han de trabajar más [...] y también a detectar las dificultades específicas de ciertos alumnos» (OEC\_C) y «son una fuente de diagnóstico muy importante para la mejora continua de los procesos de enseñanza-aprendizaje» (EDPP\_M). Dentro de esta misma dimensión, el 16% hace referencia a la «visión comparativa» que ofrecen las evaluaciones externas, «constituyen un sistema para valorar la actividad docente en relación a otros centros» (PEP\_G) y el 12% a la «visión objetiva de la realidad», en tanto es «ajena al centro y permite más objetividad y resultados claros» (EDEP\_M).

Los resultados por función principal desempeñada reflejan que en todas las categorías de esta dimensión los miembros de los equipos directivos señalan más fortalezas que el resto del profesorado: 64% frente al 34% respectivamente.

Finalmente, en los resultados de esta dimensión por etapa educativa apenas se aprecian diferencias.

### **Relacionadas con los efectos en enseñanza/centro**

Esta dimensión ha sido destacada por el 32% de la muestra total. Lo mismo que en el caso anterior, los encuestados del País Vasco son los que más fortalezas señalan (46%), seguidos por los de Cataluña (42%), Andalucía (32%), Galicia (26%) y Madrid (22%). Dentro de esta dimensión, el 15% del total de la muestra ha destacado muy especialmente la «mejora metodológica» que propician las evaluaciones externas, ya que en muchos casos exigen «otro tipo de enseñanza: más reflexiva y menos memorística» (PPP\_G) y «hacen replantear las metodologías y prioridades» (EDPP\_PV). Sin embargo, el valor dado a este aporte de las evaluaciones externas: la «mejora metodológica», muestra también diferencias por CCAA: País Vasco (24%), Cataluña (19%), Andalucía (18%), Galicia (12%) y Madrid (8%). Las demás categorías agrupadas en esta dimensión son menos mencionadas en todos los casos. Así, por ejemplo, el 6% de los encuestados destacan que las evaluaciones externas «fomentan reflexión y trabajo en equipo», en tanto «facilita un debate y análisis de la práctica docente en los diferentes departamentos [...] y una reflexión profesional muy sana para la organización educativa» (EDEC\_PV); también con diferencias destacables por CCAA, siendo más mencionada por los encuestados del País Vasco.

Como en la dimensión anterior, los resultados por función principal desempeñada muestran que los miembros de los equipos directivos señalan más fortalezas que el resto de profesores. Finalmente, en el análisis por etapa educativa apenas se aprecian diferencias, aunque es algo más apreciada en Educación Primaria, especialmente la categoría «mejora metodológica».

### **Relacionadas con las «pruebas en sí mismas»**

Las fortalezas relacionadas con esta dimensión han sido mencionadas por el 12% de la muestra total, valorando que la evaluación externa «es igual para todos y por lo tanto se puede usar como referencia» (EDPP\_G).

En los resultados por función principal desempeñada se observa que los miembros de los equipos directivos señalan algo más las fortalezas de esta dimensión que el resto de profesores y no hay diferencia destacable por etapa educativa.

### **DEBILIDADES DE LAS EVALUACIONES EXTERNAS**

La mayoría de los encuestados (97%) señala alguna debilidad de las evaluaciones externas. Solo el 3% afirma que no tienen ninguna debilidad, sin diferencias destacables por etapa educativa ni por función principal desempeñada, ni por CCAA. A continuación se presenta en la tabla 3 un resumen general de las opiniones sobre debilidades de las evaluaciones externas de aprendizaje por CCAA y muestra total; posteriormente se ofrecen los resultados por dimensión, categoría y variables contempladas.

Tabla 3  
*Opinión sobre las debilidades de las evaluaciones externas (porcentajes)*

Dimensiones	Categorías	Por CCAA					Total muestra
		A	C	G	M	PV	
		N321	N445	N239	N491	N111	N1607
Pruebas en sí mismas	Descontextualizadas	45	33	35	33	31	35
	Problemas generales de diseño	12	9	15	13	5	11
	Reducen el currículum	1	9	4	11	10	7
	Otras	11	11	10	7	11	5
	Total dimensión	69	62	64	64	57	58
Gestión evaluaciones externas	Aplicación	13	6	13	8	12	9
	Corrección	9	6	9	3	1	6
	Otros	12	7	11	8	13	10
	Total dimensión	34	19	33	19	26	25
Efectos en enseñanza-centro	Altera trabajo docente	8	8	12	22	7	13
	Efecto clasificador/ranking	5	7	8	27	9	13
	Afecta clima de centro-aula	9	8	13	19	9	12
	Otros	5	5	6	10	6	7
	Total dimensión	27	28	39	78	31	45
Resultados-información	No aportan conocimiento profundo sobre aprendizaje	5	11	10	8	10	9
	Uso propagandístico y con fines políticos	5	6	1	6	6	5
	No ofrece recomendaciones / orientaciones para mejora	2	4	5	8	10	5
	Poco valiosos	2	4	4	2	5	3
	Otros	3	3	3	6	5	5
	Total dimensión	17	28	23	30	36	27

Nota: elaboración propia. Los datos indican qué porcentaje del total de cada CCAA o muestra total ha mencionado esa categoría y dimensión. Leyenda: A=Andalucía; C=Cataluña; G=Galicia; M=Madrid; PV=País Vasco.

### Relacionadas con las «pruebas en sí mismas»

Es la dimensión más destacada, un 58% del total de los encuestados hace referencia a algún aspecto de la misma: Andalucía (69%), Galicia (64%), Madrid (64%), Cataluña (62%) y País Vasco (57%). Según la etapa educativa, hay más profesores y profesoras de Educación Primaria que

aprecian debilidades relacionadas con las «pruebas en sí mismas» que de Enseñanza Secundaria Obligatoria, 70% frente al 59%. No se aprecia ninguna diferencia por función principal desempeñada.

Dentro de esta dimensión, la debilidad más mencionada hace referencia a la «descontextualización» de las pruebas, en tanto «no tienen en cuenta singularidades propias del centro, del entorno, de las familias, de los alumnos y alumnas» (MPP\_C), señalada por el 35% de los encuestados: Andalucía 45%, Galicia 35%, Cataluña 33%, Madrid 33% y País Vasco 31%. Por otra parte, en esta categoría se aprecia una diferencia importante por etapa educativa: 42% de Educación Primaria frente al 30% de ESO. En cuanto a función principal desempeñada, los miembros del equipo directivo cuestionan algo más la descontextualización de estas pruebas que el resto del profesorado.

### **Relacionadas con los «efectos en enseñanza/centro»**

El 45% de los encuestados señala alguna debilidad relacionada con esta dimensión, de forma similar en todas las categorías, pero con diferencias por CCAA: Madrid 78%, Galicia 39%, País Vasco 31%, Cataluña 28% y Andalucía 27%. De acuerdo con los encuestados «las evaluaciones externas tienen diversos efectos que, aunque puede que sean no buscados, son negativos para el profesorado, los alumnos o el centro en general [...] para la enseñanza y para el aprendizaje» (PEP\_M).

En todas las categorías agrupadas en esta dimensión Madrid destaca sobre las demás, insistiendo sobre todo en el «efecto clasificador/ranking» (27%); mucho menos mencionada en el resto de CCAA. Algo similar sucede con la categoría «altera trabajo docente» y «afecta clima de centro-aula», señaladas por el 22% y 19% de los encuestados de Madrid respectivamente.

En esta dimensión se aprecia también una diferencia importante por función principal desempeñada, más señalada por el profesorado (53%) que por los miembros del equipo directivo (36%). Por etapa educativa, es más mencionada en ESO (50%) que EP (40%).

### **Relacionadas con «resultados/información»**

Las debilidades agrupadas en esta dimensión han sido mencionadas por el 27% de los encuestados, destacando sobre todo la opinión de que las evaluaciones externas «no aportan conocimiento profundo sobre el aprendizaje» en cuanto a «pautas o criterios de mejora en la intervención

y desarrollo del proceso de enseñanza y aprendizaje» (PPP\_M). En las respuestas de esta dimensión se aprecian también diferencias por CCAA: País Vasco 36%, Madrid 30%, Cataluña 28%, Galicia 23% y Andalucía 17%.

Por etapa educativa, es el profesorado de ESO quien encuentra más debilidades, 30% frente al 24% de los encuestados de Educación Primaria. Por función principal desempeñada las opiniones son similares, aunque el profesorado hace más referencias que los miembros del equipo directivo a la categoría referida al conocimiento sobre el aprendizaje que ofrece la información que se deriva de estas pruebas.

### **Relacionadas con la «gestión de las evaluaciones externas»**

Las debilidades relacionadas con esta dimensión han sido mencionadas por el 25% de los encuestados, destacando algo más los problemas de aplicación que los de corrección, con diferencias, como en los demás casos, por CCAA: Andalucía (34%), Galicia (33%), País Vasco (26%), Cataluña (19%) y Madrid (19%).

Por función principal desempeñada son los miembros de equipos directivos quienes señalan más debilidades de esta dimensión, 27% frente al 22% del profesorado. Finalmente, las opiniones sobre esta dimensión por etapa educativa apenas presentan diferencias.

## **ANÁLISIS Y DISCUSIÓN**

### **Sobre las fortalezas**

Dentro de las tendencias de la muestra general las fortalezas que se identifican en primer lugar coinciden con resultados previos. En las cinco CCAA se han destacado fortalezas relacionadas con los «resultados/información», al igual que en trabajos como los de Kowalski y Lasley (2009) o Spillane (2012), aunque en este caso muy especialmente en lo que se refiere a su «función diagnóstica», algo similar a lo que sostiene Skedesmo (2012), quien destaca el conocimiento que estas pruebas brindan sobre el aprendizaje. De la misma manera, se aprecia especialmente la «visión comparativa» que ofrecen; algo que ha sido señalado por Mons (2009), quien afirma que el uso comparativo de los resultados puede ser empleado para la mejora de los centros. El análisis por CCAA muestra diferencias en las opiniones sobre estas categorías que reflejan una mayor valoración en País Vasco, Cataluña y Andalucía de las mismas. Por otra parte, son mucho



más valoradas por los equipos directivos que por el resto del profesorado y más en educación primaria que en ESO.

Por otra parte, tal como se ha reflejado, aquí también se han destacado como fortaleza los «efectos en enseñanza-centro» de estas pruebas. Fundamentalmente se señala su contribución al cambio metodológico, es decir, a la mejora de la enseñanza. Se relaciona con la idea de espejo de Thélot (2002), aunque en este caso se refiere a que en muchas ocasiones estas pruebas sirven de ejemplo concreto de actividades para realizar posteriormente en el aula. Al igual que en el caso anterior, tanto la dimensión como esta categoría específica son más valoradas por los miembros del equipo directivo que por el resto del profesorado y de forma similar en Educación Primaria y en ESO. Si esto fuera así, no se daría el empobrecimiento de las prácticas docentes de las que hablan Geyer (2012) y Polesel et al. (2012), sino la mejora de las mismas, en tanto las evaluaciones externas servirían como ejemplo o fuente de actividades que se trasladan a sus propias prácticas para mejorarlas. Aunque es preciso destacar que esta valoración se hace fundamentalmente en tres de las cinco CCAA estudiadas, dato que obligaría a realizar nuevas investigaciones para buscar explicaciones sobre por qué es así en estas y no en las otras dos.

Tal como se puede apreciar, en todas las dimensiones los miembros de los equipos directivos tienen una opinión más favorable sobre las fortalezas de las evaluaciones externas que el resto del profesorado. La notable diferencia en la percepción de fortalezas por parte de unos y otros podría explicarse en tanto las evaluaciones externas podrían estar siendo un apoyo para el desarrollo de algunas de las funciones y responsabilidades que deben asumir los miembros de los equipos directivos en general y en particular los directores y directoras. Aunque no contamos con suficientes evidencias para ser concluyentes, la supuesta presión de la que hablan Feniger et al. (2015) que sufrirían los directivos de las escuelas por las demandas de las evaluaciones externas en este caso no se constataría, es decir, no serían percibidas de esta manera. Con estos datos se podría interpretar que las evaluaciones externas asumen para ellos un papel que es valorado, en tanto podrían ser útiles para hacer frente a la presión de las demandas que reciben.

Al margen de las tendencias, hay algunas singularidades a destacar que podrían dar lugar a futuras investigaciones. Por ejemplo, en el País Vasco destaca sobre las demás que estas pruebas «fomentan reflexión y trabajo en equipo»; aspecto que ha sido abordado por Darling-Hammond y Falk (2013), quienes han constatado que los procesos de reflexión sobre los resultados de las evaluaciones externas contribuyen a la mejora de la enseñanza y el aprendizaje.

## Sobre las debilidades

De la muestra total destacan las debilidades relacionadas con las «pruebas en sí mismas», aunque especialmente una de sus categorías es la más mencionada por los encuestados de todas las comunidades: la «descontextualización» de las evaluaciones externas, algo más indicado por miembros del equipo directivo que el resto del profesorado, y algo más en Educación Primaria que en ESO. En este sentido, los encuestados hacen referencia a la distancia o diferencia entre lo que las pruebas demandan y dos aspectos: a) lo que el mismo currículo establece y b) la propia realidad escolar, muy condicionada según los encuestados por su entorno inmediato.

Este aspecto, tan señalado por todos los encuestados, merece un análisis en profundidad. Aunque en las encuestas se hace referencia a un diseño descontextualizado, concretamente por los contenidos o competencias que demandan o por la poca familiarización de ciertos ejercicios o formatos de preguntas o textos; este rasgo tendría otros efectos, en tanto dicho diseño estaría condicionando los resultados obtenidos, contribuyendo así a resaltar diferencias previas y a etiquetar a estudiantes y centros educativos de acuerdo con el origen social (Bourke y Mentis, 2013; Monarca, 2012; Smith 2014; Polesel et al., 2014; Simmonds y Webb, 2013). En las respuestas de los encuestados este fenómeno también ha emergido asociado al «efecto clasificador o ranking», como práctica de jerarquización de centros a partir de listas ordenadas a partir de los resultados obtenidos en las pruebas, algo que aparece claramente destacado en la Comunidad de Madrid, con una diferencia importante con las otras CCAA analizadas.

Además de la categoría antes mencionada, dentro de los «efectos en la enseñanza/centro» también se ha señalado que las evaluaciones externas «alteran el trabajo docente», refiriéndose a aspectos coincidentes con trabajos previos: la estandarización (Geyer, 2012; Monarca, 2015; Polesel et al., 2012), la distorsión del currículo (Conley, 2015; Feniger et al., 2015; Klenowski y Wyatt-Smith, 2012), el centrar la enseñanza en estas pruebas (Hopfenbeck 2015). Al igual que en el caso anterior y lo mismo que sucede en todas las debilidades de esta dimensión, el mayor porcentaje de encuestados que la señalan son de la Comunidad de Madrid, con una diferencia destacada con las demás CCAA. Con respecto a la muestra general, algo más en el profesorado que en los miembros del equipo directivo y en ESO que en Educación Primaria.

Algo similar sucede en otra de las categorías de esta dimensión, la que hace referencia a que las evaluaciones externas «afectan al clima de centro-aula». Esta categoría condensa debilidades previamente señaladas por la investigación, como el aumento del estrés y la desmotivación de los

estudiantes (Polesel et al., 2012; Winter, 2017) o el aumento de la presión y el malestar docente en general (Wilkins, 2015).

Las otras dos dimensiones fueron bastante menos señaladas con respecto al porcentaje de respuestas de la muestra general. Sin embargo, la dimensión «resultados/información» es la segunda más mencionada en el País Vasco y en Cataluña. Dentro de esta dimensión, tanto en estas comunidades como en las demás, la debilidad más señalada hace referencia a que las evaluaciones externas «no aportan conocimiento profundo sobre el aprendizaje» o, en dos de las comunidades, que «no ofrecen recomendaciones/ orientaciones para la mejora» en el mismo sentido mencionado por otros autores (Conley, 2015; Klenowski y Wyatt-Smith, 2012; Monarca, 2012), con escasa diferencia por función principal desempeñada o por etapa educativa.

Finalmente, la dimensión que hace referencia a las debilidades relacionadas con la «gestión de las evaluaciones externas» que realiza su Comunidad Autónoma, es la menos mencionada en la muestra general, como así también en tres de las CCAA analizadas. Sin embargo, en Andalucía es la segunda debilidad que más aparece destacada y en Galicia la tercera. En ambos casos la categoría de esta dimensión señalada por un mayor porcentaje de encuestados es la referida a la «aplicación» de las evaluaciones externas, también señalada en el País Vasco. Por otro lado, se destaca como debilidad la «corrección» en Andalucía, Galicia y Cataluña, apenas mencionada en Madrid y País Vasco. En un análisis por función principal de esta dimensión, cabe mencionar que los miembros de los equipos directivos la señalan más que el resto de profesorado, muy posiblemente por el rol que cumplen en la gestión de estas pruebas, sin diferencias destacables por etapa educativa.

## CONCLUSIONES

A la luz de los datos obtenidos todo hace suponer que los sujetos construyen sus opiniones en procesos complejos multisituados atravesados por múltiples contextos y prácticas. En este sentido, puede que en ocasiones su referente sea la realidad de su propia escuela o la de su propia comunidad, mientras que en otras sea el contexto estatal. Puede que en ocasiones tome como referente sus propias vivencias sobre las evaluaciones externas, mientras que en otros se base en sus reflexiones o en los debates que circulan. Finalmente, puede que todo esto se mezcle y quede híbridamente condensado en su opinión. Esto nos obliga a tomar con cautela los resultados, pero también a admitir que pueden existir ciertas contradicciones o aparentes contradicciones que obligarían a profundizar a través de otros métodos.

En cualquier caso, más allá de las razones, se pueden extraer algunas conclusiones que emergen más o menos de forma evidente de los datos. Una de ellas es que en términos generales un mayor porcentaje de encuestados considera que las evaluaciones externas tienen más debilidades que fortalezas. Sin embargo, esta tendencia de la muestra general ofrece algunas particularidades. Por ejemplo, en la dimensión «resultados/información» los encuestados de la muestra general reconocen más fortalezas que debilidades. Por otra parte, aunque en las demás dimensiones hay un mayor porcentaje de encuestados que mencionan más debilidades que fortalezas, en un análisis por comunidad esto no siempre se mantiene. Así, en el País Vasco, Cataluña y Andalucía los encuestados mencionan más fortalezas que debilidades en la dimensión «efectos en la enseñanza/centro», siendo más destacada la diferencia en las dos primeras.

Otra conclusión general es que se señalan más debilidades y menos fortalezas en Madrid y Galicia que en el resto de las CCAA analizadas, algo que se repite en casi todas las dimensiones y categorías, dato que podría dar lugar a nuevas indagaciones.

Se puede formular otra conclusión analizando las respuestas por función principal desempeñada, que los miembros de los equipos directivos valoran más estas pruebas que el resto del profesorado. Por su parte, el análisis por etapa educativa muestra que, aunque en ambas etapas se señalan más debilidades que fortalezas, en Educación Primaria se destacan más fortalezas que debilidades en dos de las dimensiones: «efectos en enseñanza/centro» y «resultados/información».

Para finalizar es importante destacar las dos categorías que han sido señaladas por un mayor porcentaje de encuestados. Por un lado, como fortaleza, la «función diagnóstica», es decir, uno de los aspectos clave en materia de evaluación que reiteradamente se ha destacado en la literatura. Sin duda, el profesorado valora este potencial de las evaluaciones externas. Por otro lado, como debilidad, la «descontextualización», en alusión a la distancia entre lo que las pruebas demandan y lo que sucede y se hace cotidianamente en las escuelas.

Los datos ofrecidos en este artículo pueden contribuir a conocer mejor las fortalezas y debilidades de las evaluaciones externas de acuerdo con las opiniones de los miembros de equipos directos y demás profesores. Está claro que no se trata del fenómeno en sí mismo, sino de cómo es interpretado y valorado por los sujetos. En cualquier caso, estas opiniones son valiosas, en tanto surgen a partir de las propias experiencias relacionadas con las evaluaciones externas, las cuales se relacionan también con la forma en que las CCAA implementan estas evaluaciones o son usadas en el marco de

sus propios centros. Por otra parte, estas opiniones podrían tomarse como punto de partida para nuevas investigaciones.

Junto con esto, teniendo en cuenta el lugar de los sujetos en el desarrollo de las prácticas educativas, hay que tener en cuenta que las evaluaciones externas no producen efectos directos, aunque sí pueden inducir direccionalidades a los procesos de construcción de lo real: procesos de pensamiento, de reflexión, de análisis, sentimientos, relaciones; en fin, en la fabricación de instituciones y subjetividades. Sin embargo, las direccionalidades que finalmente tomen las evaluaciones externas en estos procesos dependerá de una gran variedad de aspectos: las concepciones de los sujetos implicados, otras condiciones contextuales y las características de las mismas evaluaciones externas y de las formas de gestionarlas. De esta manera, no parece apropiado establecer una conexión directa entre las pruebas y la mejora del aprendizaje (Klenowski y Wyatt-Smith 2012), como tampoco se puede anticipar que las mismas tendrán siempre un efecto negativo en la enseñanza.

## NOTAS

- 1 Este artículo se desprende de la investigación «Evaluaciones nacionales e internacionales como política de mejora de la educación en España en el contexto Europeo» financiada en el marco de la segunda convocatoria competitiva de Proyectos de Investigación Multidisciplinares de la Universidad Autónoma de Madrid (Referencia CEMU-2013-20).

## REFERENCIAS BIBLIOGRÁFICAS

- Bacon, J. (2015). The impact of standards-based reform on special education and the creation of the 'dividual. *Critical Studies in Education*, 56(3), 366-383.
- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research*, 45(2), 143-154.
- Berryhill, J., Linney, J.A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good? *International Journal of Education Policy and Leadership*, 4(5). Recuperado de <http://www.ijepl.org>
- Bishop, J. (2006). Drinking from the fountain of knowledge: Student incentive to study and learn. In A. Hanushek & F. Welsh (Eds), *Handbook of the economics of education* (pp. 909-944). Amsterdam: North-Holland.
- Bourke, R., & Mentis, M. (2013). Self-assessment as a process for inclusion. *International Journal of Inclusive Education*, 17(8), 854-867.
- Brom, L., Dalle, P., y Elbert, R. (2007). Interpretaciones sobre corrupción, democracia y desarrollo económico: entrevistas en manifestaciones colectivas de protesta. En R. Sautu (Comp.), *Práctica de la investigación cuantitativa y cualitativa: articulación entre la teoría, los métodos y las técnicas* (pp. 183-217). Buenos Aires: Lumiere.
- Cohen, L., & Manion, L. (1989). *Research methods in education*. London: Routledge.
- Conley, D. (2015). A new era for educational assessment. *Education Policy Analysis Archives*, 23(8). <http://dx.doi.org/10.14507/epaa.v23.1983>.
- Coburn, C.E., Hill, H.C., & Spillane, J.P. (2016). Alignment and Accountability in Policy Design and Implementation: The Common Core State Standards and Implementation Research. *Educational Researcher*, 45(4), 243-251.
- Darling-Hammond, L., & Falk, B. (2013). *Teacher Learning Through Assessment. How Student-Performance Assessments Can Support Teacher Learning*. Washington: Center for American progress.
- Domenech, J., Blazquez, D., de la Poza, E., & Muñoz-Miquel, A. (2015). Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain. *Educational Assessment, Evaluation and Accountability*, 27, 153-169.
- Espín, J.V. (2002). El análisis de contenido: una técnica para explorar y sistematizar información. *Educación XXI*, 4, 95-105.
- Eurydice (2009). *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Brussels: Education, Audiovisual and Culture Executive Agency.
- Falabella, A. (2014). The Performing School: The Effects of Market & Accountability Policies. *Education Policy Analysis Archives*, 22(70). <http://dx.doi.org/10.14507/epaa.v22n70.2014>
- Feniger, Y., Israeli, M., & Yehuda S. (2015). The power of numbers: the adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies and Education*, 14(2), 183-202.
- Fernández-González, N., y Monarca, H. (2018). Los sentidos de la rendición de cuentas en el discurso educativo. *Perfiles Latinoamericanos*, 26(51), 379-401.

- Flick, U. (2006). *An introduction to qualitative research*. London: SAGE Publication.
- Flórez, M.T. (2015). Systems, ideologies and history: a threedimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 3-26.
- Geyer, R. (2012). Can Complexity Move UK Policy beyond 'Evidence-Based Policy Making' and the 'Audit Culture'? Applying a 'Complexity Cascade' to Education and Health Policy. *Political Studies*, 60, 20-43.
- Harris, D., & Herrington C. (2006). Accountability, Standards and the Growing Achievement Gap: Lessons for the Past Half-century. *American Journal of Education*, 112(2), 209-239.
- Hopfenbeck, T., Flórez, M.T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: large-scale implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy & Practice*, 22(1), 44-60.
- Hsieh, H-F., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15, 1277-1288.
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65-79.
- Koch, M.J., & DeLuca, C. (2012). Rethinking validation in complex high-stakes assessment contexts. *Assessment in Education: Principles, Policy & Practice*, 19(1), 99-116.
- Komatsu, H., & Rappleye, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, 53(2), 166-191. 10.1080/03050068.2017.1300008
- Kowalski, Th., & Lasley, Th. (Eds.) (2009). *Handbook of Data-Based Decision Making in Education*. New York: Routledge.
- Ley Orgánica 2/2006, de Educación. *Boletín Oficial del Estado (España)* de 3 de mayo, 106, 17158-17207.
- Ley Orgánica 8/2013, para la Mejora de la Calidad Educativa. *Boletín Oficial del Estado (España)* de 9 de diciembre, 295, 97858-97921.
- Monarca, H. (2012). La influencia de los sistemas nacionales de evaluación en el desarrollo del currículo. *Perfiles Educativos*, 34(135), 164-176.
- Monarca, H. (2015). Sistemas de evaluación: disputas simbólicas y configuración de prácticas educativas en los procesos de globalización y reforma del Estado. En H. Monarca (Coord.), *Evaluaciones externas. Mecanismos para la configuración de representaciones y prácticas en educación* (pp. 17-42). Buenos Aires: Miño y Dávila.
- Mons, N. (2009). *Theoretical and real effects of standardised assessment*. Brussels: Eurydice.
- Parcerisa, L., y Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: Trayectoria, producción y tensiones en el sistema educativo chileno. *Education Policy Analysis Archives*, 25(89). <http://dx.doi.org/10.14507/epaa.25.3177>
- Pettersson, D., Popkewitz, T., & Lindblad, S. (2016). On the use of educational numbers: Comparative constructions of hierarchies by means of large-scale assessments. *Espacio, Tiempo y Educación*, 3(1), 177-202.
- Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The Experience of Education:*

- The impacts of high stakes testing on school students and their families.* Sydney: Whitlam Institute/University of Western Sydney.
- Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640-657.
- Rezai-Rashti, G., Segeren, A., & Martino, W. (2016). The new articulation of equity education in neoliberal times: the changing conception of social justice in Ontario. *Globalisation, Societies and Education*. 10.1080/14767724.2016.1169514
- Simmonds, M., & Webb, P.T. (2013). Accountability synopticism: How a think tank and the media developed a quasimarket for school choice in British Columbia. *The International Education Journal: Comparative Perspectives*, 12(2), 21-41.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: inconcistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23, 5-20.
- Smith, W.C. (2014). The global transformation toward testing for accountability. *Education Policy Analysis Archives*, 22(116). <http://dx.doi.org/10.14507/epaa.v22.1571>
- Spillane, J. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education*, 118, 113-141.
- Stobart, G., & Eggen, T. (2012). High-stakes testing-value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, 19(1), 1-6.
- Sun, M., Saultz, A., & Ye, Y. (2016). Federal policy and the teacher labor market: exploring the effects of NCLB school accountability on teacher turnover. *School Effectiveness and School Improvement*. 10.1080/09243453.2016.1242506
- Thélot, C. (2002). Evaluer l'Ecole. *Études*, 10(397), 323-334.
- Turnipseed, S., & Darling-Hammond, L. (2015). Accountability Is More Than a Test Score. *Education Policy Analysis Archives*, 23(11). <http://dx.doi.org/10.14507/epaa.v23.1986>
- Wilkins, C. (2015). Education reform in England: quality and equity in the performative school. *International Journal of Inclusive Education*. 10.1080/13603116.2015.1044202
- Winter, C. (2017). Curriculum policy reform in an era of technical accountability: 'fixing' curriculum, teachers and students in English schools. *Journal of Curriculum Studies*, 49(1), 55-74.



## **PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES**

Héctor Monarca. Profesor de la Universidad Autónoma de Madrid, Facultad de Formación de Profesorado y Educación, Departamento de Didáctica y Teoría de la Educación. Miembro del Grupo de Investigación sobre Políticas Educativas Supranacionales (GIPES). Líneas de investigación: política educativa, especialmente relacionadas con evaluaciones externas y desarrollo profesional docente. ORCID: <http://orcid.org/0000-0001-7749-0878>

María Fernández Agüero. Profesora de la Universidad Autónoma de Madrid, Facultad de Formación de Profesorado y Educación, Departamento de Filologías y su Didáctica. Miembro del Grupo de Investigación Análisis del Discurso y Comunicación Intercultural. Líneas de investigación: enseñanza y aprendizaje de segunda lengua, la enseñanza bilingüe y el desarrollo de la competencia intercultural en el aula.

Dirección de los autores: Universidad Autónoma de Madrid  
Facultad de Formación de Profesorado  
C/Tomás y Valiente, 3  
28049 Cantoblanco (Madrid)  
E-mail: [hector.monarca@uam.es](mailto:hector.monarca@uam.es)  
[m.fernandez@uam.es](mailto:m.fernandez@uam.es)

Fecha Recepción del Artículo: 04. Septiembre. 2017  
Fecha Modificación del Artículo: 26. Enero. 2018  
Fecha Aceptación del Artículo: 10. Febrero. 2018  
Fecha Revisión para Publicación: 22. Febrero. 2018