



## **Rendición de cuentas y política educativa: Una revisión de la evidencia internacional y futuros retos para la investigación**

*Accountability and education policy: A review of international evidence and future research challenges*

**Lluís Parcerisa y Antoni Verger**

*Departament de Sociologia, Universitat Autònoma de Barcelona*

[antoni.verger@uab.cat](mailto:antoni.verger@uab.cat) , [lluis.parcerisa@uab.cat](mailto:lluis.parcerisa@uab.cat) ,

### **RESUMEN**

La rendición de cuentas (RdC) ha devenido un tema central en la agenda educativa global, y actualmente muchos países están adoptando nuevos sistemas de RdC en sus sistemas educativos. Sin embargo, las reformas de RdC, lejos de ser homogéneas, se pueden basar en enfoques y operar en direcciones muy diferentes. Utilizando una metodología de revisión de literatura conocida como *scoping review*, este artículo analiza los principales efectos de las políticas de RdC en diversas dimensiones educativas. La evidencia revisada muestra que la RdC tiene efectos

diferentes dependiendo del diseño, la recepción y el proceso de aplicación de dicha política, así como del contexto socioeconómico e institucional de las escuelas en que se aplica. El artículo identifica además las principales lagunas de la literatura existente sobre RdC en el ámbito educativo, y sistematiza cuatro premisas con las que orientar investigación futura sobre la temática.

**PALABRAS CLAVE:** política educativa, profesionalismo docente, recontextualización, rendición de cuentas, rendición de cuentas basada en pruebas, trabajo docente.

### **ABSTRACT**

Accountability is central piece in the global education agenda, and accountability reforms are being enacted worldwide. Nevertheless, accountability reforms – far from being monolithic - can be based on very different approaches and operate in very different directions. Using a methodology based on the scoping review approach, this article aims to understand what are the main effects of accountability policies in different educational dimensions. The evidence reviewed shows that accountability policies have different effects depending on the design, the reception and the enactment process of the policy, as well as on the socioeconomic and institutional context of the schools. This article also aims to identify the main gaps in existing literature on accountability in education and systematizes four premises that can guide future research on the matter.

**KEYWORDS:** accountability, education policy, policy enactment, teachers' professionalism, teachers' work, testing-based accountability.

## 1. INTRODUCCIÓN

La rendición de cuentas (RdC) está adquiriendo cada vez más centralidad en la agenda educativa global. A pesar de que la RdC no es una nueva política en el campo educativo, las formas, el rol y los objetivos principales de la RdC han cambiado sustancialmente en los últimos años. Tradicionalmente, la RdC de carácter político y/o legal ha sido implementada con el objetivo de aumentar el control democrático y ciudadano de la educación. Sin embargo, últimamente, la RdC ha adquirido un carácter más bien gerencial (en el sentido que se basa en la obtención y evaluación de resultados escolares) y, además, ha pasado a ser concebida como una de las soluciones más importantes a los múltiples problemas y retos de los sistemas educativos contemporáneos. Actualmente, la RdC en educación es considerada como una fuente de eficiencia (es decir, una forma de promover una mejor alineación entre las aspiraciones gubernamentales y los propósitos de la escuela), de excelencia académica (una herramienta para mejorar el desempeño por parte de los centros, los profesores y los estudiantes), y de equidad (una manera de garantizar que todos los estudiantes alcancen un nivel mínimo de competencias básicas).

Las organizaciones internacionales más influyentes en el campo educativo han alimentado tan altas expectativas respecto a la RdC. En este sentido, las políticas de RdC son absolutamente centrales en las recomendaciones de los informes sobre educación elaborados por la Organización para la Cooperación y el Desarrollo Económico (OCDE), incluido el influyente *Programa para la Evaluación Internacional de Estudiantes* (PISA). De acuerdo con un reciente estudio, 29 representantes de los países de la OCDE (de 37) admitieron que las recomendaciones de PISA/OCDE sobre RdC han influido en los procesos de formulación de políticas sobre esta temática a nivel nacional (Breakspear, 2012). Para el Banco Mundial, la RdC a nivel escolar es una condición necesaria para mejorar la calidad de los maestros y el aprendizaje, ya

que permite alinear los incentivos de los diferentes agentes educativos (Banco Mundial, 2015). Por su lado, la UNESCO da mucha importancia a la RdC en su *Marco de Acción de Educación 2030*, y dedica su influyente *Informe de Seguimiento de la Educación en el Mundo* del año 2017 a este tema. Esta organización internacional considera que la introducción de mecanismos de RdC en los sistemas nacionales de educación es una condición necesaria para que los países cumplan con los objetivos de desarrollo sostenible en educación (UNESCO, 2015).

En términos de políticas, la RdC es un concepto polisémico que cubre una amplia gama de opciones y de modelos (incluyendo los modelos legal, político, burocrático y de mercado). No obstante, el modelo de RdC que actualmente está ganando más peso a nivel internacional está fundamentado en la evaluación administrativa y externa de los resultados de aprendizaje, y se centra en las escuelas y/o en los maestros como los actores que rinden cuentas. Este tipo de RdC se conoce generalmente como RdC gerencial o basada en pruebas (*testing-based accountability*) (Kamens y Benavot, 2011; Tobin, et al., 2015).

A medida que un mayor número de países adoptan sistemas de RdC - y, en particular, sistemas de RdC basados en pruebas – dichos sistemas se convierten en el foco de más investigaciones académicas. No obstante, la investigación disponible sobre la temática todavía es insuficiente para comprender en qué circunstancias y a través de qué mecanismos las políticas de RdC podrían generar los resultados esperados. De hecho, como demuestra este artículo, la evidencia sobre los efectos de las políticas de RdC sobre los resultados del aprendizaje, pero también sobre otros aspectos importantes de la educación, tales como la autonomía de los maestros, la organización de las escuelas o los procesos educativos aún no es concluyente y refleja efectos diferentes (e incluso contradictorios) en distintos lugares.

En este artículo se revisa la evidencia internacional sobre los efectos de las políticas de RdC en la educación. En concreto, sobre la base de la evidencia existente, nuestro objetivo es mostrar cómo los diferentes componentes de los sistemas de RdC (es decir, el diseño de dichos sistemas, los procesos de recontextualización [o puesta en práctica] de las políticas, las contingencias contextuales y su impacto) interactúan de forma compleja, y en qué sentido. Las preguntas que guían nuestra investigación son: 1) ¿Cuáles son los principales efectos de las políticas de rendición de cuentas en educación?; 2) ¿Cómo se manifiestan estos efectos en diferentes dimensiones educativas y cuáles son las principales áreas de acuerdo y desacuerdo?; 3) ¿Bajo qué circunstancias tienen las políticas de RdC los efectos esperados?

Con este objetivo en mente, el artículo está organizado de la siguiente manera. En la primera sección, se presentan los principales modelos de RdC en educación, y se explica por qué la RdC basada en pruebas se está convirtiendo en el modelo predominante a nivel internacional. En la segunda sección del artículo se muestra cómo la RdC basada en pruebas puede cristalizar en diseños muy diferentes de política, y por qué las decisiones sobre cuestiones de diseño son tan decisivas a la hora de entender la trayectoria de las reformas de RdC. En tercer lugar, se reflexiona sobre las diferentes dimensiones de impacto de las políticas de RdC en educación, incluyendo resultados inesperados o no deseados a nivel de escuela y de enseñanza. En cuarto lugar, nos centramos en las dinámicas y los procesos de recontextualización de las políticas de RdC - incluyendo cómo los maestros reciben, resisten y/o transforman la RdC a través de sus prácticas diarias. En quinto lugar, revisamos cómo y en qué medida los contextos institucionales y socioeconómicos en los que se aplican las políticas de RdC, también moldean la trayectoria e impacto de dichas políticas. En la sección final, en base a nuestra revisión, extraemos premisas clave y direcciones para futuras investigaciones.

Metodológicamente, el capítulo se basa en el enfoque de la revisión de alcance (o *scoping review*) de la literatura (Alegre, 2015; O'Flaherty y Phillips, 2015). Este enfoque de revisión de la literatura, en contraste con la revisión sistemática de la literatura, no parte de una pregunta de investigación muy específica ni intenta probar una teoría concreta. La *scoping review* está particularmente bien equipada para identificar las principales áreas de acuerdo y desacuerdo dentro de un determinado campo de trabajo (en nuestro caso la RdC en educación), así como las principales lagunas en el corpus actual de la literatura en este campo, en un período relativamente corto de tiempo (Mays, Roberts y Popay, 2001; Arksey y O'Malley, 2005). En este sentido, es importante subrayar que la amplitud, la profundidad y exhaustividad de la *scoping review* dependerá del propósito y/o los objetivos de la revisión. Esta metodología está compuesta por cinco fases: 1) identificación de los objetivos y de las preguntas que guían la revisión (las cuales hemos mencionado anteriormente); 2) identificación de la literatura sobre la temática; 3) selección de las investigaciones más relevantes para la revisión; 4) catalogación e interpretación de los datos; 5) ordenación, resumen y comunicación de los principales hallazgos de la revisión (Arksey y O'Malley, 2005). La búsqueda bibliográfica se realizó a partir de las bases de datos de SCOPUS y Web of Science (WoS) y se articuló en dos niveles combinando diferentes palabras clave (véase Apéndice 1): un primer nivel, relativo a la adopción y aplicación de las políticas de RdC, y el segundo nivel, relacionado con los efectos de las políticas.

La búsqueda excluyó la educación superior y se limitó al área de ciencias sociales. El límite temporal se fijó entre los años 1995 y 2014. En total, 277 documentos fueron identificados en la primera búsqueda. Seguidamente, se aplicó otro filtro relacionado durante la lectura de los resúmenes. En este caso, se descartaron aquellos estudios que no eran relevantes para el objeto de la

investigación, o bien no aportaban evidencia empírica. Durante el proceso iterativo de selección se llevaron a cabo búsquedas complementarias en otras bases de datos, como por ejemplo Google Scholar. Finalmente, 150 documentos fueron revisados, aunque no todos han sido utilizados para la elaboración de éste artículo. La información de los artículos se recolectó en fichas de lectura que incluían los siguientes ítems (adaptados a partir de Arksey y O'Malley, 2005): 1) datos de identificación: autores, año, referencia de la publicación y localización del estudio; 2) tipo de política; 3) objetivos y/o hipótesis de la investigación; 4) metodología; y 5) resultados: proceso de adopción y efectos. Por último, se ordenaron los resultados en base a las diferentes dimensiones identificadas y se llevó a cabo la síntesis de la literatura.

## **2. LA EMERGENCIA DE LA RENDICIÓN DE CUENTAS BASADA EN PRUEBAS ESTANDARIZADAS**

En términos generales, la RdC remite a los procesos, mecanismos y/o instrumentos que hacen que las organizaciones (pero también los individuos que integran las organizaciones) cumplan con sus obligaciones y se vuelvan más sensibles a sus públicos específicos (Bovens 2007, Hatch 2013). Tanto la obligatoriedad como la capacidad de condicionar la respuesta de los sujetos son características inherentes a los sistemas de RdC. De acuerdo con Bovens (2007):

Accountability is a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences (Bovens 2007, p.450).

Sin embargo, a pesar de estas características muy generales, la RdC es un concepto polisémico que cubre una amplia gama de enfoques y tipologías de políticas. Las tipologías más tradicionales de RdC son la política, la legal y la burocrática. No obstante, otros modelos de RdC presentes en determinados países o regiones son el modelo profesional (según el cual los maestros son responsables de respetar las normas profesionales y deben rendir cuentas a sus iguales –o a las organizaciones profesionales – Anderson, 2005); el modelo de mercado (que se basa en rendir cuentas a los consumidores y se promueve a través de las políticas a favor de la libertad de elección de centro – West, et al., 2011); o el modelo de RdC participativa o social, que significa que las escuelas y sus profesores son responsables ante las familias a través de la voz y el diálogo dentro de los órganos de gobierno de la escuela (West, et al., 2011). En los últimos años múltiples investigadores han desarrollado taxonomías sobre los diferentes modelos de RdC. En la Tabla 2 se presentan algunas de las principales taxonomías de la RdC en el campo educativo. Esta tabla de síntesis incorpora las principales categorías que se encuentran en la literatura (por ejemplo, la categoría de mercado, la profesional o la política), y al mismo tiempo, muestra otras categorías que a pesar de citarse bajo diferentes etiquetas, tienen un significado muy parecido (por ejemplo, en el caso de la burocrática y la jerárquica, o la contractual y la legal). La taxonomía de West et al. (2011) es la que descompone más las distintas modalidades de RdC, incorporando la modalidad en red y la participativa. Sin embargo, la taxonomía elaborada por Maroy y Voisin (2013) resulta innovadora respecto las otras, en el sentido que es la única que no se orienta en base a la naturaleza del foro, sino en la naturaleza de la obligación e incorpora la ontología de la política.



Tabla 2

*Taxonomías de la rendición de cuentas en educación*

<b>Leithwood y Earl (2000)</b>	<b>Darling-Hammod (2004)</b>	<b>West et al. (2011)</b>	<b>Maroy y Voisin (2013)</b>
<ul style="list-style-type: none"> <li>• Profesional</li> <li>• Gerencial</li> <li>• Descentralización</li> <li>• Mercado</li> </ul>	<ul style="list-style-type: none"> <li>• Política</li> <li>• Legal</li> <li>• Burocrática</li> <li>• Profesional</li> <li>• Mercado</li> </ul>	<ul style="list-style-type: none"> <li>• Professional</li> <li>• Jerárquica</li> <li>• Mercado</li> <li>• Contractual</li> <li>• En red</li> <li>• Participativa</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Strong accountability</i></li> <li>• Neo-burocrática</li> <li>• Reflexiva</li> <li>• <i>Soft accountability</i></li> </ul>

Fuente: autores

No obstante, de todos los modelos de RdC emergentes, el que tienen un alcance más global es el gerencial. La característica más distintiva de la RdC gerencial es su claro enfoque en los resultados (generalmente, resultados de aprendizaje), en lugar de poner énfasis en los *inputs* y los procesos. La aplicación de este enfoque de la RdC implica la generación de datos a través de instrumentos de evaluación estandarizada a gran escala. Es por ello que este enfoque de la RdC es también conocido como la RdC performativa (Ranson, 2003), RdC orientada a los resultados (Anderson, 2005) o RdC basada en pruebas (Hamilton, et al., 2007).

El modelo de RdC gerencial o basado en pruebas ha ganado popularidad y atracción internacional por diferentes razones. Para empezar, la mejora de los resultados de aprendizaje se ha convertido en el foco principal de las agendas de reforma educativa, tanto en los países del norte como del sur. De acuerdo a ello, la medición de los resultados de aprendizaje es una herramienta cada vez más necesaria para comprobar si las reformas logran los resultados esperados y para la promoción de la mejora escolar. En este terreno, la reforma *No Child Left Behind* adoptada en los EE.UU. en el año 2001 se ha convertido en una de las iniciativas más emblemáticas a nivel internacional. No obstante, más allá de la legislación nacional, las pruebas

estandarizadas internacionales como PISA se han convertido en un *factor* clave de las reformas de RdC basadas en pruebas. PISA no sólo ha contribuido a introducir presión competitiva a los países para mejorar los resultados de aprendizaje, sino que también se ha convertido en un referente fundamental a la hora de transferir la tecnología necesaria para adoptar pruebas de aprendizaje a nivel nacional. El surgimiento de una industria de pruebas educativas (*school testing industry*) a nivel mundial también ha presionado en una dirección similar (véase Hogan, et al., 2016).

Cuando se trata de entender el surgimiento de la RdC basada en pruebas también tenemos que hacer referencia a la emergencia de paradigmas de reforma del sector público más amplios, como la nueva gestión pública (NGP). La NGP ha contribuido a que los decisores de políticas estén más inclinados a promover estilos de gobernanza en los servicios públicos gerencialistas y orientados hacia el logro de resultados tangibles y medibles (Kalimullah, Ashraf, y Ashaduzzaman, 2012). La NGP se asocia también a corrientes de reforma educativa, tales como el movimiento de las competencias básicas comunes (*common core standards movement*), que requieren también de la implementación de sistemas de RdC basados en pruebas.

### **3. SISTEMAS DE RENDICIÓN DE CUENTAS BASADOS EN PRUEBAS DE APRENDIZAJE: MÚLTIPLES OPCIONES DE DISEÑO**

Un mismo enfoque de RdC puede adoptar formas muy diferentes en función de quién se espera que rinda cuentas, a quién se debe rendir cuentas, sobre qué se rinde cuentas, y cuáles son las consecuencias (Leithwood y Earl, 2000). En esta sección, se muestra cómo estas variables de diseño pueden interactuar y se pueden combinar de manera diferente en el marco del modelo de RdC basado en pruebas de aprendizaje.

En relación a *quién* se espera que rinda cuentas en los regímenes de RdC basados en pruebas, el *locus* se podría ubicar en el profesor individual, en grupos de profesores (por ejemplo, profesores de un mismo curso o nivel), en el director, en toda la escuela, o en redes de escuelas. El *quién* es una variable importante en el sentido de que diversos estudios muestran que, cuánto más individual sea el enfoque, más se socava la cooperación y el trabajo colegiado entre los maestros (Jones y Egley, 2004).

Los sistemas de RdC basados en pruebas de aprendizaje requieren de la presencia de una agencia de evaluación externa (que puede ser más o menos independiente del gobierno), aunque la inspección, las familias o las organizaciones profesionales pueden tener también un rol importante. Por lo que respecta a cómo se ejecuta la RdC, una variable importante es si los resultados de las evaluaciones de los centros se hacen públicos o no, y cómo (por ejemplo, si se tiene en cuenta la composición socioeconómica de la escuela o no a la hora de publicar los resultados). No obstante, la publicación de rankings escolares es una medida de RdC controvertida que, según algunos autores podría contribuir a mejorar el rendimiento académico (Boarini y Lüdemann, 2009), pero según otros podría socavar dinámicas de cooperación entre escuelas, o contribuir a la estigmatización de algunos centros (Jones y Egley, 2004).

En los regímenes de RdC basados en pruebas de aprendizaje, los resultados de los estudiantes son, casi por definición, el *objeto central* del sistema de RdC. Sin embargo, los resultados de aprendizaje se podrían combinar con la evaluación de otros componentes, como otro tipo de resultados educativos (como el nivel de abandono escolar o las tasas de graduación), los recursos y las instalaciones de la escuela, y/o los procesos educativos (a través, por ejemplo, de observaciones en el aula, revisión de portafolios docentes, etc.). En consecuencia, los sistemas de RdC podrían ser más o menos amplios por lo que respecta a los componentes

educacionales que cubren. Los diseños de RdC más comprehensivos, debido al hecho de que contemplan una gama más amplia de indicadores, podrían prevenir que las escuelas se centren en tipos específicos de resultados o pongan un excesivo énfasis en enseñar partes concretas del currículum (a costa de otras).

Por último, las *consecuencias* de las evaluaciones representan una dimensión de gran relevancia en los sistemas de RdC. En este ámbito, las opciones de diseño incluyen si la RdC es de altas consecuencias (o de alto riesgo) o no, y cuál es la naturaleza de las consecuencias: incentivos económicos, programas de formación docente, sanciones (como la intervención de la escuela, el cierre o su conversión en una escuela *chárter*), etc. Al mismo tiempo, los incentivos económicos se pueden dar a nivel de escuela o a nivel individual (en forma de bonos salariales). El importe del incentivo -si es más simbólico o más sustantivo- también condiciona el nivel de impacto en el comportamiento de la escuela o de los docentes (Escardíbul, 2015). En general, las evaluaciones de altas consecuencias ponen más presión sobre las escuelas que otras formas de evaluación, pero esto no es necesariamente algo positivo. Una presión excesiva y sin suficiente apoyo pedagógico podría promover un tipo de comportamiento oportunista o no deseado (Ohemeng y McCall-Thomas, 2013), el cual explicaremos con más detalle en la siguiente sección.

En resumen, la diversidad de opciones de política examinadas muestran que no es apropiado hablar de RdC en la educación como una política concreta. Es por ello que a la hora de realizar evaluaciones más rigurosas y precisas de los efectos de la RdC es preciso desentrañar primero las características específicas del diseño de los sistemas de RdC . En esta sección hemos dado algunos ejemplos de cómo diferentes opciones de diseño de la RdC basada en pruebas de aprendizaje podrían conllevar consecuencias distintas. A continuación, exploramos más a fondo los múltiples efectos

de la RdC en ámbitos como la organización escolar, el profesionalismo docente, los procesos educativos y los resultados de aprendizaje de los estudiantes.

#### **4. LOS MÚLTIPLES EFECTOS DE LA RENDICIÓN DE CUENTAS EN EDUCACIÓN**

Actualmente muchos países están adoptando sistemas de RdC basados en pruebas convencidos de que se trata de una buena forma de promover la excelencia académica y la mejora de los resultados del aprendizaje. Sin embargo, los sistemas de RdC también tienen efectos en otras dimensiones educativas. Algunos de estos efectos son deseados y esperados (es decir, son efectos contemplados en el diseño de la política), mientras que otros no son ni deseados ni esperados.

##### **4.1 EL CARÁCTER MIXTO DE LOS EFECTOS DE LA RENDICIÓN DE CUENTAS**

En relación al impacto en los resultados de aprendizaje, la investigación existente sobre RdC en el campo educativo muestra resultados bastante contradictorios. Varios investigadores consideran que existe una relación positiva entre las políticas de RdC y el rendimiento académico de los estudiantes, aunque muchos de ellos alertan que, por diferentes razones, esta relación puede ser bastante débil. Por ejemplo, en una investigación llevada a cabo en Israel, Rosenblatt y Shimoni (2001) encuentran que existe una asociación positiva entre la participación de los profesores de educación física en los programas de RdC y el rendimiento de los estudiantes. Sin embargo, los autores reconocen que la muestra que utilizaron era pequeña y presentaba algunas limitaciones.

Del mismo modo, Chiang (2009) aplica un análisis de regresión discontinua con datos administrativos de Florida, y llega a la conclusión de que la amenaza de sanciones a las escuelas de bajo rendimiento genera una mejora en los resultados de matemáticas de los estudiantes de primaria que persisten durante los primeros años

de la escuela media, aunque este autor señala también que dicha mejora no se pone de manifiesto en el ámbito de la lectura. Por su parte, Roderick et al. (2002), en base a datos administrativos de estudiantes de las Escuelas Públicas de Chicago, encuentran un efecto muy positivo en la mejora de las competencias básicas en dos grados (6º grado y 8º grado), aunque dicha mejora no se dio en los estudiantes de 3er grado.

Otras investigaciones son aún más explícitas acerca de la naturaleza mixta de los resultados encontrados. Este es el caso de Boarini y Lüdemann (2009), quienes concluyen que, aunque algunos aspectos de las políticas de RdC (por ejemplo, los exámenes nacionales estandarizados externos) pueden estar asociados con mejores resultados de los estudiantes, otros tienen efectos pequeños sobre los resultados de los estudiantes (por ejemplo, la publicación de los resultados de las escuelas), o no tienen. Por su parte, Ladd (1999) encuentra efectos positivos y robustos de la RdC en Dallas (Texas) para los estudiantes blancos e hispanos de séptimo grado, mientras que los estudiantes afro-americanos obtienen resultados menos positivos. Reback et al. (2014) examinan los efectos de la ley *No Child Left Behind* (NCLB) y encuentran tanto efectos positivos como efectos neutros en relación con el rendimiento académico y el disfrute en el proceso de aprendizaje. Dee y Jacob (2011) también se centran en los efectos de la ley NCLB sobre la puntuación de los estudiantes en las pruebas estandarizadas nacionales, y encuentran un efecto positivo estadísticamente significativo sobre los resultados de los estudiantes de 4º grado en matemáticas, pero no encuentran evidencia de mejora en Lectura.

Por último, otro grupo de investigadores considera que las reformas de RdC tienen efectos negativos en los resultados del aprendizaje, sobre todo desde una perspectiva de la equidad. Por ejemplo, según Andersen (2008), la aplicación de una reforma basada en la gestión del rendimiento (*performance management reform*) en las escuelas danesas tuvo efectos negativos sobre el rendimiento de los estudiantes

con bajo estatus socioeconómico (SES). Autores como Ryan (2004) señalan que el régimen de RdC basado en la prueba promovido por la ley NCLB en Estados Unidos crea incentivos que contrarían la mejora del rendimiento de los estudiantes con bajo SES. Del mismo modo, Powers (2003) muestra que la RdC basada en pruebas promovió la segregación escolar y el abandono de los estudiantes pobres y pertenecientes a minorías étnicas en California. Este tipo de RdC también puede crear incentivos para que los gobiernos reduzcan los estándares académicos y desalienta que los mejores docentes trabajen en escuelas con altos índices de población socialmente vulnerables o con bajo rendimiento, tema que retomamos en la sección final de este artículo.

La literatura revisada también contempla dimensiones de impacto que van más allá de los resultados de aprendizaje, aunque lo hace con menos frecuencia. Se trata del impacto de la RdC en, por un lado, la relación entre escuelas y, por el otro, en la identidad y el profesionalismo docente. De acuerdo a la literatura existente, las políticas de RdC pueden desencadenar tanto la colaboración como la competencia entre escuelas. En el marco de los sistemas RdC, el contexto escolar y el papel del director parecen mediar estratégicamente sobre si el comportamiento de las escuelas deviene más cooperativo o más competitivo. Se trata de un tema de gran relevancia dados los beneficios potenciales que las dinámicas de cooperación pueden generar en los sistemas educativos (Sahlberg, 2010; Keddie, 2015), pero que todavía ha sido poco explorado.

En cuanto a la identidad profesional y a la autonomía profesional de los docentes, la investigación existente -por lo general de carácter etnográfico- apunta que las políticas de RdC basadas en el desempeño tienen más desventajas que beneficios. Estas políticas tienden a limitar la autonomía profesional de los docentes, a generar mayores niveles de estrés, y/o a erosionar las relaciones sociales dentro de la

escuela (ver Day, 2002; Jeffrey y Woods, 1996; Maxcy, 2009). A pesar de las implicaciones obvias de estos tipos de efectos en los niveles de motivación docente, también hay investigaciones que contemplan que la RdC tiene efectos positivos en la motivación del profesorado (Finnigan, 2010).

#### 4.2 EFECTOS NO DESEADOS

Son cada vez más numerosos los trabajos de investigación que reflejan los efectos no deseados de las políticas de RdC en educación, especialmente de las políticas de RdC de altas consecuencias y que promueven dinámicas competitivas. En ciertos contextos, estos tipos de RdC pueden alentar la emergencia de comportamientos oportunistas por parte de escuelas, maestros y directores. El concepto de comportamiento *oportunista* se refiere a una amplia gama de prácticas, incluyendo la selección de estudiantes (Jennings, 2010), la enseñanza orientada a la prueba o *teaching to the test* (Ohemeng y Thomas-McCall, 2013), el triaje educativo (Booher-Jennings, 2005; Perryman, et al., 2011; Reback, 2008), e incluso la realización de trampas en el momento de tomar la prueba (Jacob y Levitt, 2003). Los actores educativos adoptan este tipo de comportamientos con el fin de maximizar sus resultados en las pruebas. Estos comportamientos emergen con más fuerza entre aquellas escuelas y/o aquellos docentes sometidos a mayores niveles de presión externa para lograr mejores resultados educativos, sobre todo debido a la amenaza de sanciones que estos sistemas implican en el caso de tener bajo rendimiento.

Un tipo de comportamiento oportunista identificado por la literatura es el llamado triaje educativo que, en palabras de Ladd y Lauen (2010, p.429), se refiere a "*triaging out' students well below and well above grade level and 'triaging in' students close to grade level*". Estos autores, en su análisis de los efectos distributivos de diferentes programas de RdC en Carolina del Norte, no encuentran evidencia de la



existencia de la práctica del triaje. Sin embargo, sus resultados contrastan con los de otro estudio realizado en Texas, el cual sí demuestra que los docentes responden a los incentivos para mejorar los resultados centrándose en el grupo de estudiantes que más se aproxima a la tasa de aprobación, a expensas de otros grupos de estudiantes (Booher-Jennings, 2005 ).

Ohemeng y McCall-Thomas (2013) llevaron a cabo un estudio sobre los comportamientos no deseados generados por las pruebas estandarizadas en Ontario. En esta investigación se llevaron a cabo entrevistas con los diferentes agentes educativos clave con el fin de comprender las presiones e incentivos que emanan de un sistema de pruebas estandarizadas. Muchos de los profesores entrevistados argumentaron que la presión para obtener buenos resultados en las pruebas los condujo a adoptar prácticas de *teaching to the test* y a centrarse en aquellas áreas del currículum que serían evaluadas en las pruebas (Ohemeng y Thomas McCall, 2013). Del mismo modo, Jones y Egle (2004) encontraron que una de las quejas más frecuentes de los profesores en relación con el sistema de pruebas de altas consecuencias fue la necesidad de pasar gran parte de su tiempo preparando a sus estudiantes para realizar la prueba de forma efectiva –a través de la enseñanza de contenidos similares a los incluidos en la prueba y de la realización de simulacros de pruebas. Estos hallazgos son consistentes con los resultados de otros estudios (véase Au, 2007; Hamilton, et al., 2007; Linn, 2000; Menken, 2006; Volante, 2004). Se trata sin duda de prácticas que resultan preocupantes desde el punto de vista educativo ya que conllevan una reducción del currículum y socavan la idea de la enseñanza y el aprendizaje como un proceso creativo y comprensivo.

## **5. EL PROCESO DE ADOPCIÓN Y DE RECONTEXTUALIZACIÓN DE LAS POLÍTICAS DE RENDICIÓN DE CUENTAS**

La adopción de nuevas formas de RdC en educación acostumbra a ser un proceso político que genera algún nivel de conflicto y resistencia, lo cual acaba revirtiendo en la calidad de la implementación de dichas políticas y en sus resultados finales.

Estudiar la trayectoria de las reformas educativas y cómo dicha trayectoria condiciona los resultados de las reformas no es una tarea sencilla. Según McDermott (2007), en toda reforma educativa acostumbra a haber mucha diferencia entre, por un lado, los objetivos o las aspiraciones declaradas de la reforma y, por el otro, las prácticas que se llevan a cabo en la realidad en nombre de dicha reforma. Que una política educativa cuente con un buen diseño y sea promovida con buenas intenciones es una condición necesaria pero no suficiente para que la política sea exitosa. En general, la percepción y la adhesión de los agentes educativos hacia la política en cuestión resulta clave a la hora de comprender los éxitos y fracasos de dicha política. Esto es especialmente relevante en el caso de la adopción de sistemas de RdC ya que se trata de reformas que acostumbran a generar controversia y respuestas diferenciales por parte del profesorado, incluyendo estrategias de negociación y de resistencia tanto abiertas como encubiertas.

Según Ball et al. (2012), en la literatura sobre política educativa, el concepto de implementación de políticas es utilizado a menudo de forma restringida y lineal, ya que omite los procesos (micro)políticos que se dan en las escuelas. Estos autores sugieren que los textos y los programas de políticas no son implementados de manera mecánica, sino que son interpretados y recontextualizados de forma activa por parte de los actores educativos a nivel micro-institucional de formas a menudo poco

previsibles. Es por ello que en lugar del concepto de implementación, Ball et al. (2012) sugieren utilizar el concepto de “*policy enactment*”, el cual sirve para capturar y comprender el proceso contingente, y su vez creativo, de decodificación y recodificación mediante el cual las políticas se ponen en acción en las escuelas.

Por lo tanto, para tener una visión comprensiva y global de las reformas educativas y entender mejor sus efectos es necesario abrir la caja negra que hay entre el diseño y los resultados de las políticas, y analizar aquellos procesos de carácter político y/o semiótico mediante los que las políticas se traducen en prácticas concretas. Esto significa, en el caso que nos ocupa, analizar el proceso mediante el cual las reformas de RdC han sido adoptadas y recibidas por parte de los actores educativos encargados de aplicarlas. A continuación, exploramos el rol de los docentes en tanto que intérpretes, traductores y productores de las reformas de RdC.

## **5.1 EL PERSONAL DOCENTE ANTE LAS POLÍTICAS DE RENDICIÓN DE CUENTAS**

La complicidad de los diferentes actores que operan a nivel escolar (en particular los directores de centro y los docentes) es una condición necesaria para garantizar que la política sea aplicada con éxito. No obstante, para que una reforma disponga de la complicidad de los actores es necesario que estos perciban que la reforma tendrá un impacto positivo y que la consideren significativa para su desarrollo profesional y para la mejora escolar.

En un estudio comparativo desarrollado a nivel europeo, Müller y Hernández (2010) exploraron las percepciones de los docentes sobre los sistemas de RdC. En esta investigación, aplicaron una encuesta a una muestra de 1.100 docentes de cada país (Finlandia, Suecia e Irlanda), 44 entrevistas basadas en historias de vida con un total de 22 docentes, observación y grupos de discusión. Sus resultados muestran que

en Europa predomina el escepticismo y la desconfianza entre los docentes respecto a los sistemas de RdC (en particular, sobre los sistemas de RdC basados en el desempeño o en pruebas). Al mismo tiempo, su estudio señala el riesgo de que los sistemas de RdC se conviertan en un mecanismo vacío que sólo produzca una presión negativa, y que no consiga motivar a los docentes. Para evitar que la RdC se convierta en algo superfluo o en una influencia negativa para la mejora escolar, estos investigadores sugieren que los mecanismos de RdC deben orientarse a reforzar la profesionalización docente y responder a las necesidades reales de la comunidad educativa (Müller y Hernández, 2010).

De manera similar, autores como Jones y Egley (2004) exploraron las percepciones de los docentes sobre el sistema de exámenes de altas consecuencias a través de una encuesta aplicada a 708 docentes del estado de Florida, según la cual la mayoría de los docentes (79,9%) creía que las políticas de RdC implementadas en Florida no iban en la buena dirección. En base a dicho resultado, Jones y Egley (2004) concluyen que estas percepciones negativas condicionan en gran medida las respuestas de los docentes al sistema de RdC y condenan la reforma al fracaso.

Las políticas de RdC basadas en el desempeño acostumbran a alterar las relaciones de poder existentes en los sistemas de gobernanza escolar (Maxcy, 2009). Estas políticas tienden a otorgar mayores cotas de poder a la administración, los directores o las familias, a la vez que socavan la capacidad de decisión del profesorado. Cuando los docentes perciben que las políticas les desempoderan o restringen su autonomía profesional (por ejemplo, sobre decisiones en el ámbito curricular), estos generalmente adoptan estrategias de resistencia con la que hacer frente a la aplicación de las reformas. Por ejemplo, en Sudáfrica los docentes boicotearon la aplicación de un nuevo sistema de RdC ausentándose de las reuniones,

negándose a enviar sus registros o a realizar el trabajo estipulado en el marco del programa de RdC (Mosoge y Pilane, 2014).

Los docentes también pueden adoptar estrategias intermedias de negociación y evasión (Perryman, et al., 2011). Por ejemplo, Palmer y Rangel (2011) encontraron que los docentes de Texas intentaron equilibrar la atención de las necesidades de sus alumnos con las presiones procedentes del sistema de RdC para orientar la enseñanza a los contenidos de las pruebas. Por otro lado, Kostogriz y Doecke (2011) muestran cómo los docentes de Australia fueron capaces de reivindicar su estatus profesional y su autonomía bajo un sistema de RdC, mediante una praxis que, en cierta manera, desafiaba el marco regulatorio de las reformas basadas en estándares. Otras investigaciones han encontrado resultados similares, y han observado cómo los docentes negocian su agencia profesional en el marco de las reformas de RdC y desafían de manera creativa la restricción de su autonomía profesional (véase Hardy, 2014; Osborn, 2006; Robinson, 2012).

En definitiva, las respuestas de los docentes frente a los sistemas de RdC pueden incluir un amplio abanico de reacciones que van desde el conflicto y la resistencia abierta, hasta estrategias de resistencia encubierta y el consentimiento. Dichas respuestas, en lugar de darse por supuesto, deberían conformar un eje fundamental en la investigación empírica sobre los efectos de la RdC.

## **6. EL CONTEXTO SOCIAL Y ECONÓMICO DE LAS REFORMAS DE RENDICIÓN DE CUENTAS**

Los contextos institucionales y socio-económicos donde las políticas de RdC son aplicadas condicionan los resultados de dichas políticas. La puesta en práctica de las políticas de RdC no tiene lugar en el vacío sino que se produce en escuelas

concretas ubicadas en contextos con importantes condicionantes externos, sociales y materiales. En palabras de Ball et al. (2012, p. 19):

Policies enter different resource environments; schools have particular histories, buildings and infrastructures, staffing profiles, leadership experiences, budgetary situations and teaching and learning challenges.

Así pues, más allá del rol de los actores a título individual, para comprender los efectos de las reformas de RdC deben tenerse en cuenta las interrelaciones entre las dimensiones materiales, sociales y estructurales que configuran las dinámicas de las escuelas, las cuales pueden ejercer presión y actuar como facilitadores o limitadores en el despliegue de las políticas educativas (Ball, et al., 2012).

Muy a menudo, la investigación sobre procesos de mejora y reforma escolar y, en particular, sobre reformas de RdC omite las variables de naturaleza contextual o institucional de las escuelas. Sin embargo, también encontramos excepciones. Por ejemplo, Keddie (2014) realizó un estudio de caso cualitativo en una escuela inglesa ubicada en un contexto socialmente aventajado. Según esta autora, los factores situados, profesionales, materiales y externos mediaron estratégicamente en la puesta en práctica de las políticas de autonomía escolar y de RdC. Entre otras cosas, estos factores posibilitaron a la escuela la adopción de dinámicas cooperativas y le permitieron desafiar la “cultura de la auditoria” así como algunos de los efectos negativos que acostumbran a derivar de ésta. Por otro lado, en base a un estudio conducido en Sudáfrica, Douglas (2005) concluye que los recortes en la plantilla docente, las ratios (alumnos-profesor) excesivamente grandes, y los materiales inadecuados para la enseñanza y el aprendizaje constriñeron la puesta en práctica de la reforma del sistema de evaluación del desempeño.

La comparación entre escuelas con diferente rendimiento académico es una manera habitual de intentar capturar la influencia de las variables contextuales y el rol de los factores institucionales en la puesta en práctica de las reformas. En general, la evidencia existente sobre esta temática muestra que las escuelas que, según el sistema de RdC, se encuentran en período de prueba (*on probation*) tienden a adoptar estrategias a corto plazo (más “cosméticas”) para salir de su situación, mientras que las escuelas con alto rendimiento, que no tienen tantas presiones externas, tienden a tener más posibilidades de introducir mejoras en el ámbito de la enseñanza y en el ámbito instructivo. En base a la evidencia recogida en un estudio etnográfico realizado en Chicago, Diamond y Spillane (2004) muestran que el estatus (período de prueba o alto rendimiento) de la escuela influye en la adopción de respuestas diferenciales frente a las presiones procedentes del sistema de RdC. Los hallazgos de esta investigación muestran que las escuelas en período de prueba adoptaron respuestas más superficiales y más reactivas a las sanciones y a las presiones externas, mientras que las escuelas de alto rendimiento alinearon sus respuestas con las aspiraciones reales de los *policy-makers* (por ejemplo, introduciendo cambios en la instrucción e innovaciones). Por supuesto, los factores institucionales y socioeconómicos tienden a estar fuertemente relacionados. En su análisis comparativo de las escuelas de Chicago, Finnigan (2010) observa que los factores socioeconómicos están directamente relacionados con las escuelas catalogadas como “en período de prueba”, ya que un 92% de las escuelas de Chicago con dicho estatus tienen una mayoría de estudiantes de bajos ingresos.

Las reformas educativas basadas en políticas de RdC de altas consecuencias tienden a favorecer que los docentes más motivados se planteen dejar las escuelas en período de prueba, exacerbando así las desigualdades entre escuelas. En una investigación conducida en once escuelas en período de prueba de Estados Unidos,

Mintrop (2003) encontró que los maestros se mostraban escépticos frente a las políticas de RdC y consideraban que el sistema de RdC no reconocía suficientemente su esfuerzo en un contexto educativo con grandes dificultades. Por ello, este autor concluye que el sistema de RdC promovía que los docentes más motivados se plantearan abandonar las escuelas en período de prueba para ir a trabajar a escuelas más aventajadas.

Por su lado, Falabella (2014) examina la influencia del contexto institucional de las escuelas chilenas frente al sistema de RdC. Esta investigadora distingue tres grupos de escuelas (escuelas de alto rendimiento, escuelas de rendimiento medio-bajo, escuelas de bajo rendimiento). En el caso del primer grupo, la autora observa que las políticas de RdC sirven para reforzar el prestigio social de las escuelas y su posición aventajada en el mercado educativo local. Sin embargo, las políticas de RdC no generan incentivos para introducir cambios pedagógicos o de gestión en las escuelas aventajadas. El segundo grupo de escuelas dispone de una demanda muy fiel, normalmente por razones más simbólicas que académicas, de manera que dichas escuelas tampoco tienen incentivos para introducir mejoras educativas en base a los resultados de las pruebas estandarizadas. Finalmente, el último grupo de escuelas, formado por las escuelas de bajo rendimiento, es muy sensible a los resultados del sistema de RdC. Generalmente se trata de escuelas con elevados porcentajes de alumnado vulnerable donde la RdC genera una sensación de estrés debido a la vigilancia y a las sanciones que se asocian a los malos resultados (Falabella, 2014).

## **7. DISCUSIÓN FINAL: CUATRO PREMISAS PARA FUTURAS INVESTIGACIONES**

Actualmente numerosos países están adoptando políticas de RdC basadas en pruebas de aprendizaje como una forma de mejorar la eficacia de sus sistemas



educativos y la productividad de su profesorado, así como de garantizar unos niveles mínimos de equidad educativa. No obstante, esta forma de RdC se está globalizando a pesar de que todavía hay muchas incertidumbres sobre si ésta funciona o no, en qué sentido y bajo qué circunstancias.

En este artículo, hemos revisado la literatura existente sobre el tema a través del método de la *scoping review*. La adopción de esta metodología nos ha permitido indagar en el amplio espectro de investigación sobre el tema y sistematizar sus resultados más relevantes. Es más, este método nos ha permitido reflexionar sobre cómo los diferentes diseños de los sistemas de RdC, procesos de recontextualización, y contingencias de carácter contextual e institucional pueden conducir a efectos y resultados muy diferenciados. De nuestra revisión se derivan cuatro premisas básicas para la investigación futura sobre la temática. Éstas son:

- 1) La RdC no es una política educativa específica, más bien abarca opciones de política muy diferentes; diferentes diseños de RdC pueden dar lugar a resultados divergentes, pero también a un amplio abanico de respuestas y procesos de “*enactment*” o recontextualización.
- 2) A pesar de que la RdC en educación es vista principalmente como una herramienta para mejorar el aprendizaje de los estudiantes, ésta también tiene efectos en otras dimensiones educativas. Algunos de estos efectos son deseados y/o esperados, mientras que otros no lo son.
- 3) La RdC es generalmente una opción política que genera conflicto y controversia entre la comunidad educativa; el proceso de adopción de sistemas de RdC -así como las características particulares de estos sistemas- condicionan su recepción y puesta en práctica por parte de los docentes, así como sus consecuencias finales.

- 4) La RdC en educación, como cualquier otra política global, no es resiliente al contexto. Los contextos institucionales y socioeconómicos en los que se promulgan y aplican las políticas de RdC, condicionan la trayectoria y los resultados finales de estas políticas.

Estas cuatro premisas pueden contribuir a sentar las bases de investigaciones sobre RdC más comprehensivas y que interpelen a la realidad de diferentes contextos educativos. Al fin y al cabo, estas cuatro premisas nos indican cuáles son los elementos que aún están poco investigados en la literatura existente sobre la RdC en educación, y apuntan a problemas de causalidad en el estudio de los efectos de este tipo de reformas.

En concreto, la premisa 1 refleja la importancia de examinar las características específicas del diseño de las políticas de RdC como un primer paso necesario para promover evaluaciones más rigurosas de las trayectorias y efectos de dichas políticas. Debido a sus importantes implicaciones, las opciones de diseño no deberían ser vistas simplemente como decisiones técnicas, sino como decisiones con implicaciones profundamente políticas. Como hemos visto, algunas de estas opciones tienen la capacidad de alterar las relaciones de poder fundamentales dentro de los sistemas educativos y generar tipos de respuestas muy diferentes entre los docentes y otros actores educativos clave.

De la premisa 2 se deriva que los efectos no esperados/no deseados de la RdC deben ser también contemplados en los modelos de análisis de estudios futuros. Todavía hay una parte importante de la investigación sobre la RdC en educación que se centra casi exclusivamente en los resultados de aprendizaje como la única variable dependiente. Sin duda, esto es algo que restringe profundamente nuestra comprensión de la naturaleza sistémica de los efectos de la RdC en el campo

educativo. Por ello, cabe indagar todavía más sobre los efectos de las medidas de RdC sobre el trabajo y los comportamientos de los docentes, y acerca de qué tipo de sistemas de RdC son más proclives a fomentar actitudes de colaboración (o de competencia) entre los agentes educativos.

La premisa 3 remite a la importancia de abrir la caja de negra que existe entre el diseño de las políticas y sus resultados finales. La implementación de políticas no es un proceso mecánico, sino que un proceso moldeado fuertemente por variables subjetivas (motivación docente, identidades profesionales, percepciones y creencias, múltiples interpretaciones, etc.) que son difíciles de medir en términos únicamente cuantitativos. A pesar de la existencia de un cuerpo emergente de investigación que se centra en este tipo de variables, la investigación que indaga sobre las conexiones entre, por un lado, la respuesta de maestros, directores y otros agentes educativos ante las políticas de RdC y, por otro lado, diferentes tipos de resultados educativos es todavía escasa.

Por último, la premisa 4 refleja la importancia de considerar variables de carácter contextual a la hora de analizar el despliegue y el impacto de las políticas de RdC. A través de nuestra revisión, hemos detectado que la mayoría de investigaciones sobre el tema están sesgadas hacia un enfoque de la eficacia escolar (o *school effectiveness*) que por lo general no presta suficiente atención a cómo el contexto socioeconómico interfiere en cómo las políticas de RdC operan a nivel escolar. La realización de estudios comparativos entre países, estados y/o regiones podría contribuir a poner de relieve la compleja interacción entre entornos institucionales y socioeconómicos, procesos de recontextualización y los efectos de la RdC en múltiples dimensiones educativas, que pueden ir desde el rendimiento académico a la satisfacción y el desarrollo profesional del personal docente.

En general, la forma en que los profesores son concebidos y tratados en el marco de los sistemas de RdC, pero también cómo los docentes interpretan, interactúan y actúan en el marco de estos sistemas es clave para entender los resultados diferenciales - y en ocasiones contradictorios - de la RdC en el campo educativo. La RdC de altas consecuencias tiende a generar más resistencia que otros modelos de RdC, entre otras cosas, debido a los problemas de atribución que plantea cuando trata de definir qué parte del éxito y del fracaso estudiantil es atribuible a los docentes. Autores como Leithwood y Earl (2000) señalan que los sistemas de RdC serían más legítimos si no responsabilizaran a los docentes de forma casi exclusiva de los resultados de aprendizaje de sus alumnos –ya que estos resultados se explican por muchos factores. En lugar de ello, los sistemas de RdC deberían más bien responsabilizar al profesorado de hacer un uso de los recursos disponibles lo más beneficioso posible de cara a avanzar hacia un amplio abanico de objetivos educativos que incluyan objetivos de rendimiento académico, pero también de equidad y de gobernanza democrática de la educación.

## REFERENCIAS BIBLIOGRÁFICAS

- Alegre, M. A. (2015). *Què funciona en educació: la pregunta necessària*. Barcelona: Fundació Jaume Bofill & Ivalua. Recuperado de: [http://www.ivalua.cat/documents/1/07\\_07\\_2015\\_15\\_54\\_14\\_Què\\_funciona\\_educacio.pdf](http://www.ivalua.cat/documents/1/07_07_2015_15_54_14_Què_funciona_educacio.pdf)
- Arksey, H., y O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. doi: <http://dx.doi.org/10.1080/1364557032000119616>

- Andersen, S. C. (2008). The impact of public management reforms on student performance in Danish schools. *Public Administration*, 86(2), 541-558. doi: 10.1111/j.1467-9299.2008.00717.x
- Anderson, J. A. (2005). *Accountability in education*. Paris: UNESCO-International Institute for Educational Planning.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267. doi: 10.3102/0013189X07306523
- Ball, S. J., Maguire, M., y Braun, A. (2012). *How schools do policy. Policy Enactment in Secondary Schools*. New York: Routledge.
- Boarini, R., y Lüdemann, E. (2009). The role of teacher compensation and selected accountability policies for learning outcomes: An empirical analysis for OECD countries. *OECD Journal: Economic Studies*, 1, 1-20. doi:10.1787/eco\_studies-v2009-art10-en
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42(2), 231-268. doi: 10.3102/00028312042002231
- Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. *European Law Journal*, 13(4), 447-468. doi: 10.1111/j.1468-0386.2007.00378.x
- Breakspear, S. (2012). The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. OECD Education Working Papers, 71. OECD Publishing. doi: <http://dx.doi.org/10.1787/5k9fdfqffr28-en>

- Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93(9), 1045-1057.  
doi:10.1016/j.jpubeco.2009.06.002
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *The Teachers College Record*, 106( 6), 1047-1085.
- Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, 37(8), 677-692.  
doi:10.1016/S0883-0355(03)00065-X
- Dee, T. S., y Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.  
doi:10.1002/pam.20586
- Diamond, J., y Spillane, J. (2004). High-stakes accountability in urban elementary schools: challenging or reproducing inequality?. *The Teachers College Record*, 106(6), 1145-1176.
- Douglas J. (2005). *Teacher professionalism and education transformation*. Braamfontein: Centre for Education Policy Development (CEPD).
- Escardíbul, O. (2015). Is the introduction of pay-for-performance salary incentives for teachers linked to students' academic performance advisable? *What Works in Education Series*, 1, Barcelona: Fundació Jaume Bofill & Ivalua. Recuperado de:  
[http://www.ivalua.cat/documents/1/27\\_10\\_2015\\_13\\_44\\_00\\_WhatWorksinEducation\\_english.pdf](http://www.ivalua.cat/documents/1/27_10_2015_13_44_00_WhatWorksinEducation_english.pdf)
- Falabella, A. (2014). "Do national test scores and quality labels trigger school self-assessment and accountability? A critical analysis in the Chilean context."

*British Journal of Sociology of Education*, 37(5), 1-18.

doi:10.1080/01425692.2014.976698

Finnigan, K.S. (2010). Principal Leadership and Teacher Motivation under High-Stakes Accountability Policies, *Leadership and Policy in Schools*, 9(2), 161-189.

doi:10.1080/15700760903216174

Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., y Robyn, A. (2007).

*Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Rand Corporation.

Hardy, I. (2014). A logic of appropriation: enacting national testing (NAPLAN) in Australia, *Journal of Education Policy*, 29(1), 1-18. doi:

10.1080/02680939.2013.782425

Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14 (2), 113-138

Hogan, A., Sellar, S. y Lingard, B. (2016). "Corporate Social Responsibility and Neo-Social Accountability in Education: The Case of Pearson plc." In Verger, A., Lubienski, C., & Steiner-Khamsi, G. (eds). *World Yearbook of Education: The Global Education Industry* (pp. 107-124). New York: Routledge.

Jacob, B. A. y Levitt, S.D. (2003). Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating, *Quarterly Journal of Economics*, 118(3), 843-877. doi: 10.1162/00335530360698441

Jeffrey, B., y Woods, P. (1996). Feeling deprofessionalised: The social construction of emotions during an OFSTED inspection. *Cambridge Journal of Education*, 26(3), 325-343. doi: 10.1080/0305764960260303

- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247. doi: 10.1177/0038040710375688
- Jones, B. D., y Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39), 1-34. doi: <http://dx.doi.org/10.14507/epaa.v12n39.2004>
- Kalimullah, N. A., Ashraf, K. M., y Ashaduzzaman, M. N. (2012). New public management: Emergence and principles. *BUP Journal*, 1(1), 1-22.
- Keddie, A. (2014). 'It's like Spiderman... with great power comes great responsibility': school autonomy, school context and the audit culture. *School Leadership & Management*, 34(5), 502-517. doi: 10.1080/13632434.2014.938040
- Keddie, A. (2015). School autonomy, accountability and collaboration: a critical review. *Journal of Educational Administration and History*, 47(1), 1-17. doi: 10.1080/00220620.2015.974146
- Kostogriz, A., y Doecke, B. (2011). Standards-based accountability: Reification, responsibility and the ethical subject. *Teaching Education*, 22(4), 397-412.
- Ladd, H. F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, 18(1), 1-16.
- Ladd, H. F., y Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, 29(3), 426-450.



- Leithwood, K., y Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education*, 75 (4), 1-18.
- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.  
doi: 10.3102/0013189X029002004
- Maroy, C., y Voisin, A. (2013). As transformações recentes das políticas de accountability na educação: desafios e incidências das ferramentas de ação pública. *Educação & Sociedade*, 34(124), 881-901.
- Maxcy, B. (2009). New public management and district reform: Managerialism and deflection of local leadership in a Texas school district. *Urban Education*, 44(5), 489-521. doi: 10.1177/0042085908318778
- Mays, N., Roberts, E. y Popay, J. (2001) "Synthesising research evidence". In N. Fulop, N., Allen, P., Clarke, A., & Black, N. (eds.) *Studying the Organisation and Delivery of Health Services: Research Methods*. (pp. 188-220). London: Routledge.
- McDermott, K.A. (2007). "Expanding the Moral Community" or "Blaming the Victim"? The Politics of State Education Accountability Policy. *American Educational Research Journal*, 44(1), 77–111. doi: 10.3102/0002831206299010
- Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521–46.
- Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 10(3), 1-30.

- Mosoge, M. J., y Pilane, M. W. (2014). Performance management: the neglected imperative of accountability systems in education. *South African Journal of Education*, 34(1), 1-18. doi: 10.15700/201412120947
- Müller, J. y Hernández, F. (2010). On the geography of accountability Comparative analysis of teachers' experiences across seven European Countries. *Journal of Educational Change*, 11(4), 307-322.
- O'Flaherty, J., y Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, 27, 85-95. doi: <http://dx.doi.org/10.1016/j.iheduc.2015.02.002>
- Ohemeng, F., y McCall-Thomas, E. (2013). Performance management and “undesirable” organizational behaviour: Standardized testing in Ontario schools. *Canadian Public Administration*, 56 (3), 456-477. doi: 10.1111/capa.12030
- Osborn, M. (2006). Changing the context of teachers' work and professional development: A European perspective. *International journal of educational research*, 45(4), 242-253. doi:10.1016/j.ijer.2007.02.008
- Palmer, D., y Rangel, V. S. (2011). High stakes accountability and policy implementation: Teacher decision making in bilingual classrooms in Texas. *Educational Policy*, 25(4), 614-647. doi: 10.1177/0895904810374848
- Perryman, J., Ball, J.S., Maguire, M., y Braun, A. (2011). Life in the Pressure Cooker – School League Tables and English and Mathematics Teachers' Responses to Accountability in a Results-Driven Era, *British Journal of Educational Studies*, 59(2), 179-195. doi: 10.1080/00071005.2011.578568

- Powers, J. M. (2003). An analysis of performance-based accountability: Factors shaping school performance in two urban school districts. *Educational Policy*, 17(5), 558-585. doi: 10.1177/0895904803256789
- Ranson, S. (2003). Public accountability in the age of neo-liberal governance, *Journal of Education Policy*, 18(5), 459-480. doi: 10.1080/0268093032000124848
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5), 1394-1415. doi: 10.1016/j.jpubeco.2007.05.003
- Reback, R., Rockoff, J., y Schwartz, H. L. (2014). Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind. *American Economic Journal: Economic Policy*, 6(3), 207-241. doi: 10.3386/w16745
- Robinson, S. (2012). Constructing teacher agency in response to the constraints of education policy: adoption and adaptation. *Curriculum Journal*, 23(2), 231-245. doi: 10.1080/09585176.2012.678702
- Roderick, M., Jacob, B. A., y Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24 (4), 333-357.
- Rosenblatt, Z., y Shimoni, O. (2001). Teacher accountability: An experimental field study. *Journal of Personnel Evaluation in Education*, 15 (4), 309-328.
- Ryan, J. E. (2004). Perverse Incentives of the No Child Left behind Act, *The New York University Law Review*, 79, 932-989.

Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11 (1), 45-61.

Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., y Nyamkhuu, T. (2015). *Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region*. Melbourne: Australian Council for Educational Research (ACER).

UNESCO. (2015). *Framework for Action Education 2030: Towards inclusive and equitable quality education and lifelong learning for all*. Recuperado de: <http://www.uis.unesco.org/Education/Documents/wef-framework-for-action.pdf>

Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*, 35. Recuperado de: <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>

West , A., Mattei, P., y Roberts, J. (2011). Accountability and Sanctions in English Schools, *British Journal of Educational Studies*, 59 (1), 41-62. ISSN 0007-1005

World Bank. (2015). *What Matters Most for School Autonomy and Accountability: A Framework Paper*. SABER Working Paper Series, 9. Recuperado de: [https://olc.worldbank.org/sites/default/files/Whatmatters\\_SAA\\_framework\\_paper\\_Demas\\_2014\\_4.pdf](https://olc.worldbank.org/sites/default/files/Whatmatters_SAA_framework_paper_Demas_2014_4.pdf)

## APÉNDICE 1

Tabla 1

*Sintaxis*

```
TITLE-ABS-KEY ("education"OR "education reform"OR "School")AND TITLE-ABS-KEY("Effective school research"OR "Efficiency"OR "Effectiveness"OR "Effects" OR "Outcomes")AND TITLE-ABS-KEY("New Public Management"OR "NPM"OR "School autonomy"OR "Institutional autonomy"OR "Accountability"OR "Merit-based pay" OR "School-based management"OR "Public management reform") AND NOT ("Higher education") AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"ARTS" ) ) AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) )
```

```
TITLE-ABS-KEY ("education"OR "education reform"OR "School")AND TITLE-ABS-KEY("Enactment"OR "Implementation")AND TITLE-ABS-KEY("New Public Management"OR "NPM"OR "School autonomy"OR "Institutional autonomy"OR "Accountability"OR "Merit-based pay" OR "School-based management"OR "Public management reform") AND NOT ("Higher education") AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"ARTS" ) ) AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) )
```

Fuente: autores

## AGRADECIMIENTOS

Este trabajo ha sido realizado en el marco del programa de doctorado en Sociología de la Universidad Autónoma de Barcelona y ha contado con el apoyo de la beca FPU del Ministerio de Educación, Cultura y Deporte español (ref. FPU13/01090). Asimismo, los autores también agradecen el financiamiento entregado por el European Research Council bajo el programa marco Horizon 2020 de investigación e innovación de la Unión Europea, número de subvención 680172 - REFORMED: <http://reformedproject.eu>.