

CAPÍTULO 2

POLÍTICAS DE EVALUACIÓN ESCOLAR Y RENDICIÓN DE CUENTAS: SU IMPACTO EDUCATIVO EN UN SISTEMA DE BAJAS CONSECUENCIAS

Mauro C. Moschetti¹

Gerard Ferrer-Esteban²

Edgar Quilabert³

Antoni Verger⁴

DOI: 10.52695/978-65-5456-163-1.2

Introducción

En las últimas décadas, numerosos países han puesto en marcha reformas para modernizar sus sistemas educativos y mejorar simultáneamente su efectividad y equidad. La centralidad creciente de la educación en la llamada sociedad del conocimiento, los acelerados cambios tecnológicos y la globalización, así como el impacto de evaluaciones internacionales como PISA o PIRLS en el debate público, han consolidado un énfasis de la

1 Universitat Autònoma de Barcelona. E-mail: mauro.moschetti@uab.cat.

2 Universitat Oberta de Catalunya.

3 Universitat Autònoma de Barcelona.

4 Universitat Autònoma de Barcelona.

política educativa en elevar el rendimiento académico del alumnado. Bajo esta lógica, y con la intención de superar las rigideces asociadas a los modelos burocráticos del siglo XX, se han impulsado políticas que promueven la autonomía de los centros, introducen evaluaciones basadas en resultados académicos y refuerzan los mecanismos de rendición de cuentas. Estas estrategias buscan que las escuelas amplíen sus márgenes de decisión sobre aspectos organizativos y pedagógicos, desarrollen proyectos de mejora educativa propios y asuman una responsabilidad más directa sobre su desempeño.

Actualmente, la autonomía escolar, la evaluación y la rendición de cuentas (RdC) se articulan como un bloque de políticas interdependientes. A cambio de una mayor libertad organizativa, presupuestaria o curricular, las administraciones educativas exigen mayores niveles de supervisión sustentada en evidencias — como pruebas externas estandarizadas o procesos de inspección. Esta lógica permite que docentes y equipos directivos adapten sus proyectos educativos a las características del entorno y accedan a un abanico más amplio de recursos formativos y herramientas orientadas a la mejora escolar. Sin embargo, esta mayor autonomía implica también una mayor responsabilidad sobre los resultados obtenidos.

Aunque esta tendencia de reforma educativa se ha generalizado en países muy diversos — en lo que algunos autores describen como un proceso de convergencia internacional (Breakspear, 2012; Teltemann; Jude, 2019) —, la evidencia sobre su eficacia es mixta (Torres, 2021). En algunos contextos, la tríada autonomía, evaluación y RdC fomenta la innovación, refuerza la enseñanza en el aula y eleva el rendimiento escolar (De Grauwe, 2005; Greany, 2018); en otros, favorece la competencia entre escuelas, la segmentación de la oferta y prácticas centradas en *teaching to the test* (una enseñanza centrada en la preparación para las pruebas), la reducción del currículo a los contenidos estrictamente evaluados o la separación del alumnado por niveles de rendimiento, con efectos negativos sobre la equidad (Ferrer-Esteban, 2015; Pagès *et al.*, 2023; Werler; Færevaaag, 2017).

Por último, estas políticas persiguen objetivos distintos según el contexto: mientras que en algunos países se orientan a mejorar la calidad y la equidad educativas, en otros buscan fomentar la elección escolar y la

competencia entre centros (Verger; Fontdevila; Parcerisa, 2019). Su impacto, evolución y configuración dependen en gran medida de las características de los sistemas educativos en los que se implementan, así como del diseño específico de las propias políticas. La mayoría de los estudios sobre estas políticas se han centrado en países anglosajones, donde han predominado modelos de evaluación con altas consecuencias. En estos contextos, se ha observado una intensificación de comportamientos oportunistas y la aparición de efectos no deseados (Mittleman; Jennings, 2018).

En Cataluña (España), foco de este capítulo, el sistema de evaluación y RdC se caracteriza por ser eminentemente *low-stakes*, es decir, de bajas consecuencias, en el sentido de que los resultados no se hacen públicos ni determinan la carrera profesional del profesorado ni la financiación de los centros educativos. La institucionalización de estas políticas se produjo con la *Llei d'Educació de Catalunya* (2009) y sus decretos reglamentarios (2010), que reforzaron la autonomía de los centros, reglamentaron las evaluaciones externas y profesionalizaron el liderazgo y la gestión escolar. Dicha ley aspiraba a fomentar la innovación pedagógica y una gestión más flexible en un sistema históricamente muy centralizado.

En este contexto, las pruebas de Competencias Básicas (CB), iniciadas en el curso 2000-2001, han ido adquiriendo un peso creciente como instrumento de diagnóstico y mejora. Sin embargo, aún existe un número limitado de estudios que analicen cómo se implementan estas políticas en la práctica en contextos *low-stakes*, qué significados adquieren para los distintos actores escolares y qué usos efectivos se hacen de los datos generados por las pruebas en el ámbito de los centros educativos.

Este capítulo tiene por objetivo analizar los efectos y significados de las políticas de evaluación externa y rendición de cuentas a partir de un estudio de caso en Cataluña. Lejos de tratarse de un análisis meramente local, el caso catalán permite observar las tensiones, mediaciones y resignificaciones que emergen entre los marcos normativos de la política educativa y su implementación cotidiana en los centros escolares. Cataluña ha desarrollado un modelo que combina autonomía escolar con evaluaciones de carácter fundamentalmente formativo, en un entorno cada vez más competitivo — marcado por una notable presencia del sector privado,

el descenso demográfico y la ampliación de la elección escolar. En este contexto, el estudio ofrece claves analíticas útiles para comprender cómo estas políticas operan realmente en la práctica y en qué condiciones generan determinados efectos, tanto esperados como no esperados.

El capítulo se organiza del siguiente modo. Primero presentamos el marco conceptual del estudio, centrado en la relación entre autonomía escolar y evaluación externa, y exponemos el enfoque teórico que orienta nuestro análisis, basado en las lógicas de apropiación de las políticas educativas por parte del profesorado. A continuación, describimos el contexto normativo catalán en el que se insertan estas políticas. En el apartado siguiente detallamos el enfoque metodológico, que combina un cuestionario administrado a una muestra representativa de 78 centros de enseñanza obligatoria con entrevistas semiestructuradas a docentes y equipos directivos de 20 centros.

Posteriormente, presentamos los resultados empíricos, organizados en tres apartados: en el primero analizamos las fuentes de información que utilizan los centros para orientar sus procesos de mejora; en el segundo, examinamos el uso que se hace de los datos generados por las evaluaciones externas, en particular por las pruebas de competencias básicas; en el tercero, abordamos las experiencias de presión asociadas a dichas pruebas y sus efectos en la cultura escolar; finalmente, cerramos el capítulo con una discusión de los hallazgos y algunas recomendaciones de política derivadas del análisis. El caso de Cataluña se propone como un ejemplo paradigmático de sistema *low-stakes*, cuya trayectoria ofrece claves valiosas para repensar las políticas de evaluación en otros contextos con características similares.

Diseño y efectos de las políticas de evaluación y rendición de cuentas escolar

Las políticas de evaluación y rendición de cuentas han ido habitualmente de la mano de procesos de descentralización y autonomía escolar, pues exigen que un amplio abanico de agentes — desde la administración hasta las familias y los centros — asuman responsabilidad sobre los resultados educativos. Estas políticas se apoyan en estándares de aprendizaje

y marcos de competencias, sistemas de monitorización, planes de mejora e incentivos vinculados a los logros académicos. En la práctica, son sobre todo los docentes y los equipos directivos quienes rinden cuentas de los avances de sus estudiantes, medidos por evaluaciones externas que funcionan como indicadores de calidad a nivel de aula, escuela, zona e incluso del sistema en su conjunto (Nichols; Harris, 2016).

En las dos últimas décadas ha crecido exponencialmente el número de países que aplican evaluaciones o pruebas externas censales, sobre todo en lengua, ciencias y matemáticas (UNESCO, 2019; Verger *et al.*, 2019). Aunque puedan diferir en nivel educativo, frecuencia, materias y modalidad de administración, las pruebas persiguen, en mayor o menor medida, dos objetivos: por un lado, ofrecer información diagnóstica y formativa que oriente al profesorado, ajuste el currículo y ayude a planificar el desarrollo profesional; por otro lado, reforzar la rendición de cuentas mediante la publicación de resultados y la vinculación de incentivos o sanciones a los logros alcanzados (Clarke *et al.*, 2021).

Diversos estudios han identificado también beneficios potenciales asociados a estos sistemas. Se ha señalado, por ejemplo, que las pruebas externas pueden elevar las expectativas de rendimiento de estudiantes y centros (Dee y Jacob, 2011), facilitar la identificación de buenas prácticas y orientar los esfuerzos hacia estrategias respaldadas por evidencia (Datnow *et al.*, 2007; Vogler, 2017), dirigir recursos adicionales a los grupos con mayores dificultades y promover una cultura de transparencia y mejora continua entre el profesorado. Según la OCDE, los efectos más positivos de la rendición de cuentas tienden a observarse cuando la publicación de resultados se combina con altos niveles de autonomía en la gestión de recursos humanos y económicos (OCDE, 2011; Torres, 2021). Sin embargo, en el contexto europeo esta relación aparece empíricamente más débil y heterogénea (European Commission, 2018), lo que sugiere que los resultados dependen de cómo se articulan estos instrumentos en cada sistema educativo.

Una variable central para comprender los efectos diferenciales de la rendición de cuentas es el tipo y nivel de consecuencias que se derivan de las evaluaciones externas. Cuando las pruebas están asociadas a incentivos o sanciones significativas para estudiantes, docentes o centros

— lo que caracteriza a los sistemas *high-stakes* —, abundan las evidencias de efectos no deseados: la enseñanza se concentra en los contenidos evaluados, se estrecha el currículo y aumenta el estrés entre docentes y alumnado; además, se incrementan las prácticas de selección y agrupación que comprometen la equidad interna de los centros (Jerrim y Sims, 2021; Mittleman y Jennings, 2018). A ello se suma que la combinación de autonomía escolar y publicación de resultados puede potenciar la competencia entre escuelas: las que obtienen mejores puntuaciones las utilizan como recurso para atraer alumnado con mayor capital económico y cultural, mientras las familias más informadas eligen a partir de estas referencias, contribuyendo así a una mayor segmentación del sistema (Hamilton *et al.*, 2002; Allen y Vignoles, 2007).

En contraste, los sistemas de bajas consecuencias conciben las evaluaciones externas como herramientas principalmente formativas. En estos casos, los resultados sólo generan implicaciones para los centros con rendimientos persistentemente bajos y, cuando esto ocurre, suelen traducirse en mayor acompañamiento institucional y oportunidades de desarrollo profesional, en lugar de sanciones. El uso de los datos se orienta a promover la reflexión pedagógica y la identificación de áreas de mejora. Aun así, diversos estudios muestran que, incluso en estos entornos, el profesorado puede cuestionar o resistir estas políticas cuando las percibe como mecanismos de control externo que interfieren en su autonomía profesional (Maroy y Voisin, 2017; Barbana *et al.*, 2020; Thiel *et al.*, 2017).

Asimismo, distintos autores han subrayado que la efectividad de la rendición de cuentas no depende únicamente de los instrumentos utilizados, sino del uso que se haga de los datos generados. Para que esta información contribuya realmente a la mejora, resulta imprescindible que los actores implicados la perciban como válida y dispongan de la formación, el tiempo y los recursos necesarios para analizarla e interpretarla de manera rigurosa. De lo contrario, la infrautilización o el uso indiscriminado de los resultados puede convertir la inversión en evaluaciones externas en una estrategia ineficaz e incluso costosa (Terhart, 2013; UNESCO, 2019).

En cualquier caso, los estudios empíricos sobre los efectos de las políticas de evaluación y rendición de cuentas *low-stakes* siguen siendo

mucho más limitados que aquellos centrados en contextos *high-stakes* (Levatino *et al.*, 2024).

Profesorado y apropiación de políticas educativas

El análisis de la implementación de las políticas educativas de gobernanza escolar plantea importantes retos. Parte de su complejidad radica en que se trata de políticas en continua evolución, replanteamiento y ajuste. Además, su impacto sobre el comportamiento y las prácticas del profesorado suele ser indirecto y está fuertemente mediado por una multiplicidad de factores, como la cultura escolar, las dinámicas organizativas internas, las trayectorias profesionales o las condiciones materiales de los centros.

En este estudio, centrado en el caso de Cataluña, analizamos la implementación de las políticas de evaluación y RdC en el ámbito escolar. También exploramos su impacto en la práctica educativa y en la organización interna de los centros educativos. De forma general, el estudio busca examinar hasta qué punto se materializan en la práctica los principales supuestos y efectos esperados de las políticas de autonomía, evaluación y RdC, especialmente en lo que respecta a los procesos de mejora e innovación educativa y a la conformación de equipos docentes cohesionados. Por lo tanto, centramos nuestro análisis en la asociación de estas políticas con las prácticas pedagógicas y organizativas, así como al uso que hacen los centros escolares de las evaluaciones externas, entre otras fuentes de información, para producir cambios y mejoras escolares.

En este estudio entendemos que la asociación de las políticas de autonomía y evaluación, a nivel de sistema, con las prácticas escolares concretas, a nivel de centro, no depende únicamente de cómo la administración educativa regula e implementa las políticas. Es una asociación que también está mediada por una serie de variables relevantes de contexto, como la composición social y el origen del alumnado, el carácter público o privado de los centros, sus resultados en las evaluaciones de las competencias básicas o las percepciones y opiniones que los docentes y equipos directivos tienen tanto de estas políticas como de su propia práctica profesional. Por tanto, los análisis de este estudio se llevan a cabo considerando la diversidad de los centros en titularidad, los niveles

de rendimiento, la composición y vulnerabilidad social del alumnado y las etapas educativas.

Para profundizar en la importancia de las variables de tipo contextual, en los análisis también consideramos la posición de las escuelas en su mercado educativo local (MEL) que, según la definimos en este estudio, viene determinada por tres factores: la reputación de los centros en su entorno más inmediato, su rendimiento académico medio en las pruebas de CB y el nivel de demanda que reciben. Se trata de una variable que, a pesar de que en investigación educativa se utiliza menos que otras variables de contexto educativo — como la composición social del alumnado o el estilo de liderazgo de la escuela —, nos ayuda a entender por qué los centros escolares experimentan de forma diferente y responden con distintas actuaciones ante el marco regulativo y la presión competitiva (Maroy y van Zanten, 2009).

En relación con esto último, creemos que la adhesión del personal docente a este tipo de políticas — cómo las viven y las interpretan — es muy variable y tiene implicaciones directas en la implementación de la política educativa (Malen, 2006). Una de las principales premisas teóricas de este estudio es que los procesos de significación de las políticas por parte del profesorado — lo que en la literatura internacional se conoce como *'enactment'* — es clave para entender cómo éstas se traducen en diferentes decisiones y prácticas en los centros educativos (Ball *et al.*, 2011; Spillane *et al.*, 2002). En esta línea, el concepto de *'apropiación'* que desarrollan Levinson *et al.* (2009) es particularmente útil para entender cómo una política es interpretada de forma activa — y a menudo creativa — por una determinada comunidad de práctica, responsable de su implementación en función de las condiciones institucionales y las normas culturales y sociales que la definen.

La adopción de políticas de autonomía, evaluación y RdC en el sistema educativo catalán

La *Llei d'Educació de Catalunya* (LEC) de 2009 marcó un hito incorporando, por un lado, disposiciones de la Ley Orgánica de Educación española

(LOE) (2006) y trazando, por otro, las bases de un modelo educativo propio en la comunidad autónoma (Farré, 2009). Inspirada en los preceptos de la Nueva Gestión Pública, favoreció la descentralización, la planificación estratégica y la gestión por resultados, lineamientos plasmados en 2010 en los decretos que reglamentarían la Autonomía de Centro, la Dirección Escolar y la Evaluación Educativa. En materia de autonomía, las escuelas ampliaron sus competencias de gestión económica y organizativa, asumieron la definición de estructuras y proyectos y profesionalizaron la figura del director, tanto mediante formación y procesos de selección específicos como a través de una mayor capacidad para gestionar las plantillas docentes.

A diferencia de los decretos de autonomía y dirección, el decreto de evaluación tuvo un recorrido limitado y, debido a la crisis financiera, en 2011 se suspendió la creación de una agencia de evaluación profesionalizada e independiente. En el momento de realizar este estudio, el Consejo Superior de Evaluación del Sistema Educativo (CSASE) — dependiente del Departamento de Educación y con recursos económicos y humanos restringidos — era el organismo encargado de administrar las pruebas de Competencias Básicas (CB) y de coordinar la participación de Cataluña en evaluaciones internacionales. Según la Resolución EDU/3.614/2022, las pruebas de CB constituyen “una de las líneas de mejora de la educación en Cataluña”, ya que proporcionan información global sobre el sistema y permiten que cada centro valore sus propios resultados y planifique acciones de mejora. Aunque la LEC no permite la difusión de los resultados con fines promocionales o comparativos, los datos generados alimentan otros instrumentos de planificación escolar y de supervisión por parte de la inspección educativa.

Por su parte, la profesionalización de las direcciones de centro ha liberado a la Inspección Educativa de tareas de gestión para centrarse en la evaluación y el acompañamiento. Hoy, la normativa vigente refuerza la autonomía de gestión de los equipos directivos y la responsabilidad por los resultados de aprendizaje ante familias, inspección y administración. Se espera que la devolución de datos genere debate en los equipos docentes y directivos e impulse planes de mejora, ya sea de forma autónoma o con la ayuda de agentes como la Inspección, los llamados Centros de Recursos Pedagógicos o la llamada Red de Competencias Básicas.

Este modelo de gobernanza escolar sitúa la autonomía escolar en el núcleo del sistema — y como condición necesaria para la articulación de proyectos educativos innovadores y contextualizados. Las evaluaciones externas, por su parte, constituyen un componente clave para diagnosticar el rendimiento del alumnado y orientar la gestión escolar. También se contemplan como uno de los instrumentos para la promoción docente por méritos, aunque esta vía ha tenido hasta ahora un desarrollo limitado en comparación con otras modalidades de promoción de carácter más burocrático.

En este estudio, analizamos estos supuestos, así como la percepción y el grado de penetración de estas políticas en los centros, y su alineación con los objetivos normativos que las sustentan.

Métodos y datos

Nuestro estudio se basa en un enfoque de métodos mixtos que combina un cuestionario administrado a una muestra representativa de 78 centros de enseñanza obligatoria de Cataluña (155 directoras/es y 852 docentes) con entrevistas semiestructuradas a directores/as y profesorado de 20 centros seleccionados por heterogeneidad (titularidad, complejidad y composición socioeconómica). De forma complementaria, se emplearon fuentes secundarias del Departamento de Educación de la Generalitat de Catalunya para caracterizar variables contextuales (categoría de complejidad, porcentaje de alumnado extranjero, tamaño del claustro y número de estudiantes), así como entrevistas a informadores clave del sistema educativo (representantes del Departamento, del CSASE y de la Inspección Educativa) para triangular resultados.

En la fase cuantitativa, se utilizó un muestreo estratificado (por área urbana, titularidad y nivel educativo), combinado con un muestreo aleatorio proporcional al tamaño (PPS) y la selección aleatoria de 10 docentes por centro, considerando criterios de carga lectiva y asignatura. Se realizaron análisis descriptivos (distribuciones, correlaciones, comparaciones de medias) y modelos de regresión lineal y logística. Las variables de interés incluyeron índices de autonomía educativa y de gestión, presión hacia el rendimiento, cultura de pruebas y posición en el mercado educativo local,

mientras que los controles abarcaron características de escuela (titularidad, composición socioeconómica, rendimiento medio, reputación y tamaño) y de profesor (edad, género, experiencia, contrato, afiliación sindical, uso de datos), con efectos fijos de provincia. La imputación de las variables con valores perdidos se hizo con *missing dummies* (Cohen; Cohen, 1983).

La fase cualitativa, basada en estudios de caso, incluyó entrevistas semiestructuradas de 45–60 minutos de duración a 20 directoras/es y 48 profesoras/es, complementadas con análisis documental y notas de campo. El guion abordó tanto aspectos biográficos como el contexto escolar, las percepciones sobre evaluación y rendición de cuentas, las respuestas organizativas y pedagógicas, la autonomía en el proyecto educativo y las estrategias frente a la competición escolar. Todas las entrevistas se transcribieron íntegramente y se codificaron mediante una estrategia flexible que combinó códigos deductivos y emergentes, agrupándolos temáticamente y analizando sus coocurrencias para identificar patrones. El análisis documental de proyectos institucionales, planes de mejora y webs de los centros reforzó la interpretación de los hallazgos.

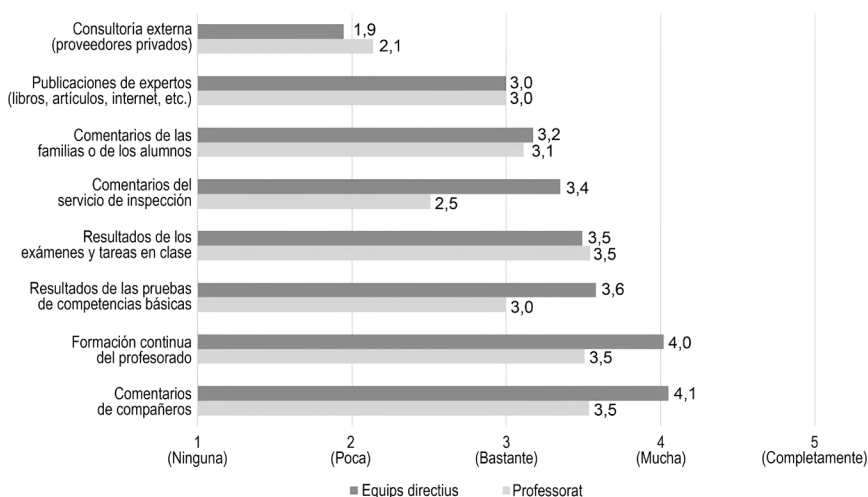
Resultados

Organizamos los resultados del estudio en torno a los siguientes ejes: en primer lugar, analizamos las fuentes de información y evaluación utilizadas por el profesorado y los equipos directivos, así como el uso y la importancia que les otorgan, con especial énfasis en las pruebas externas. A continuación, examinamos la presión que el sistema de evaluación y rendición de cuentas ejerce sobre los centros educativos catalanes, y cómo ésta influye en sus decisiones organizativas y prácticas pedagógicas. Aunque el diseño de la política catalana responde a un modelo formalmente *low-stakes*, los resultados muestran que su implementación no siempre se traduce en experiencias de baja presión percibida, ni en usos de los datos predominantemente formativos. Los resultados permiten observar cómo los centros utilizan las evaluaciones externas junto con otras fuentes de información, qué factores influyen en su uso y qué prácticas emergen en contextos con mayor nivel de presión o con mayores limitaciones para trabajar con los datos.

Fuentes de datos para la mejora

Los datos de las pruebas externas constituyen solo una de las varias fuentes a las que recurren los actores escolares para detectar áreas de mejora. Los docentes y directivos valoran especialmente los comentarios entre compañeros, la formación continua y el análisis de los resultados de exámenes y tareas en el aula, ya que les aportan información práctica y orientadora (gráfico 1). Las direcciones conceden además una alta utilidad a los resultados de las pruebas de competencias básicas y a los informes de la inspección; en cambio, el profesorado tiende a valorar menos estas fuentes externas, salvo cuando se traducen en discusiones y reflexiones concretas sobre los resultados de su propia práctica evaluativa.

Gráfico 1 – Utilidad de las fuentes de información para mejorar la calidad de la enseñanza del centro, equipos directivos y docentes



Fuente: Proyecto Reformed (Cataluña).

El hecho de que los equipos directivos otorguen mayor relevancia a los insumos de la inspección educativa responde a su canal de interlocución casi exclusiva con este servicio. Sin embargo, este vínculo varía notablemente según el inspector/a asignado al centro: en unos centros se percibe un estilo distante y centrado en la supervisión de documentos e indicadores, mientras que en otros predomina un acompañamiento cercano, orientado

a asesorar en los procesos de mejora pedagógica. Esta diversidad de enfoques modula la utilidad real que cada escuela extrae de los informes y recomendaciones de la inspección. Así lo ilustra una directora:

Lo cierto es que yo, por ejemplo, he visto tres inspecciones diferentes en esta escuela en cinco años: la primera era una inspección muy cercana, que valoraba muchísimo todo el trabajo que hacíamos, nos ayudaba a recabar recursos, incluso nos acompañaba en el aula y nos observaba. O sea, una inspección espectacular. Luego tuvimos otra inspección que estuvo un año, que hubo como una transición. Ahora tenemos una inspección que ha aparecido en dos ocasiones durante todo el curso. Una de ellas era por los indicadores del proyecto de dirección, que deben evaluarse y no sé qué. Mucho desde la burocracia, la verdad. Y es un acompañamiento muy técnico, con una mirada menos comprensiva, podríamos decir, de las dificultades del día a día (Directora 20, escuela pública 20).

Sin embargo, también hay quien considera que, de forma progresiva, se está transitando de un modelo de supervisión más distante a un modelo de acompañamiento más cercano, tal y como se ilustra en la siguiente entrevista:

La imagen histórica que se tiene de la Inspección de que vienen aquí a joderte, esto ha cambiado, ellos también tienen otro rol, ahora el inspector viene a aquí a ayudar, a informarse, viene a trasladar información a gente que está por encima de él, no es lo mismo. Debo decir también que cuando nos hicieron la auditoría pedagógica, ellos fueron también muy cercanos, muy humanos, no teníamos la sensación de estar haciendo [una rendición] de cuentas, sino que era más como un acompañamiento, un asesoramiento, [...] debemos decir que estamos contentos (Directora 7, escuela pública 7).

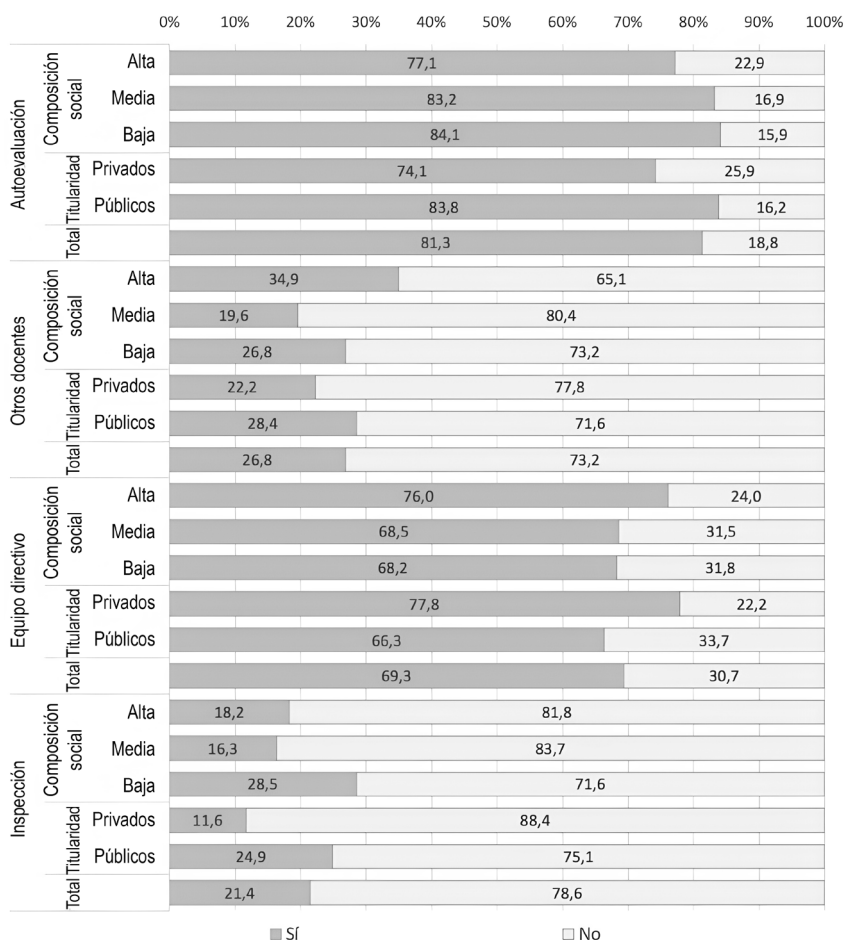
Por lo general, la inspección educativa no actúa de forma intervencionista en lo pedagógico: rara vez entra en el aula u observa directamente la implementación del proyecto educativo. Su relación con el profesorado suele ser indirecta, a través del equipo directivo, y se centra en revisar la documentación de planificación y organización del centro, así como en analizar indicadores de resultados y mejora generados internamente y por la administración. Esta dinámica quedó registrada en las entrevistas realizadas a directoras y maestros:

P: ¿Qué rol tiene la inspección en este caso? R: Ella cada cierto tiempo se reúne con el equipo directivo y debe supervisar el trabajo que se está realizando y, a final de curso, el equipo directivo debe exponer, con el departamento, inspección, etc., los resultados alcanzados. Hay unos indicadores muy claros con los niños, hay que recabar información, pruebas, para ser un poco más objetivo, a veces cuestan ponderar estas cosas [...] Yo creo que la inspección delega más en el equipo directivo, no sé si quizás porque asumen muchas escuelas, no sé, es el equipo directivo quien hace esta función (Docente 39).

Junto a las fuentes de información para la mejora educativa, también exploramos qué agentes participan en la evaluación de la labor docente (gráfico 2). La modalidad más habitual es la autoevaluación, utilizada por el 81% del profesorado, que resulta especialmente frecuente en la escuela pública, mientras que en los centros concertados (de titularidad privada y con financiamiento público) y aquellos de menor complejidad su práctica disminuye. En segundo lugar, los equipos directivos desempeñan un rol relevante para casi el 70% de los docentes, con un peso mayor en la escuela concertada (78%) que en la pública (66%). Por el contrario, la evaluación entre iguales y la realizada por la inspección son poco habituales: solo el 27% identifica a compañeros como evaluadores y el 21% recurre a la inspección, pese a su función oficial de supervisión. Aunque la inspección evalúa algo más en la pública (25%) que en la concertada (11%), su presencia es limitada y tiende a concentrarse en los centros de menor composición social.

En esta primera sección observamos que la autonomía en el uso de información no conduce necesariamente a usos profundos o formativos de los datos. La evaluación externa constituye una referencia más dentro de un ecosistema de fuentes, no siempre la más influyente, lo que anticipa las tensiones entre autonomía, valoración de las pruebas y uso real para la mejora que emergen en los apartados posteriores.

Gráfico 2 – Fuentes de evaluación docente: ¿quién evalúa tu trabajo?



Nota: los porcentajes se han calculado considerando el profesorado que identificaba al menos una fuente de evaluación

Fuente: Proyecto Reformed (Cataluña).

Uso de los datos de las evaluaciones externas

Los centros pueden utilizar los datos de las pruebas estandarizadas para múltiples fines: evaluar el trabajo docente, detectar estudiantes con necesidades educativas, ajustar el currículo, informar a familias y profesorado, agrupar alumnos por nivel, compararse con otros centros o reforzar su reputación. Su empleo más habitual consiste, sin embargo, en definir y monitorizar el plan de mejora escolar. De hecho, los resultados de las pruebas de competencias básicas suelen servir de referencia en planificaciones anuales, memorias, proyectos de dirección y acuerdos de corresponsabilidad (gráfico 3).

Este uso genérico de los datos como herramienta de ‘mejora escolar’ explica que se declare con más frecuencia que otros usos más específicos, e incluso controvertidos. Las entrevistas muestran que, en la práctica, los datos de las pruebas suelen generar debate y reflexividad en el claustro docente y actúan como motor de los procesos de mejora. Esto se expresa en el siguiente testimonio de una directora:

P: ¿Las pruebas de competencias han contribuido a mejorar la calidad educativa del centro y el rendimiento de los estudiantes? R: A ver, te hace reflexionar para realizar mejoras. El hecho de que como equipo directivo te propongas llegar a la media o intentar superarla, evidentemente hace poner en marcha un engranaje de reflexión, que puedes estar más de acuerdo o menos de acuerdo con las pruebas en sí, pero sí es verdad que dices: “como esto es lo que tenemos” debemos intentar sacarle el máximo provecho. Dices: “tienes esto, diseña unos objetivos hacia la mejora” y al final es a partir del resultado de competencias básicas (Directora 1, escuela pública 1).

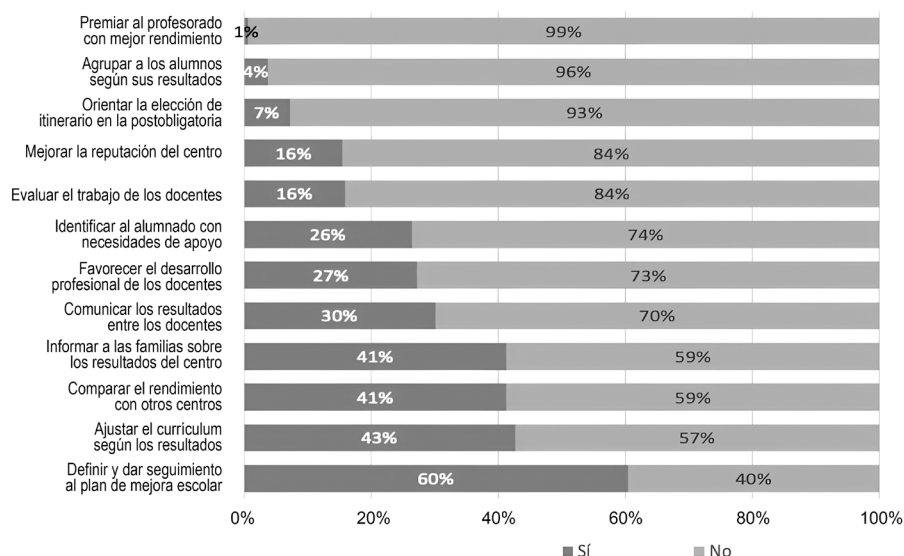
El resto de usos de los datos tiene un papel más secundario en el contexto educativo catalán. Poco más del 40% del profesorado recurre a ellos para ajustar el currículo, compararse con otras escuelas o informar a las

familias sobre el nivel de consecución de aprendizajes (gráfico 3). Entre el 25% y el 30% los emplea para identificar alumnado con necesidades de apoyo o impulsar el desarrollo profesional.

Los usos más instrumentales, como agrupar a los estudiantes según su rendimiento, premiar a los de mejor rendimiento o realzar la reputación del centro, son muy poco frecuentes en Cataluña, como también lo es el uso de estos datos para detectar a quienes requieren refuerzo. Sin embargo, el bajo aprovechamiento formativo puede estar relacionado con los plazos de administración de las pruebas de competencias básicas, que se aplican al final de cada etapa (primaria y secundaria) y cuyos resultados se retornan a los centros educativos cuando el alumnado evaluado ya se ha graduado o está a punto de hacerlo. Dos directoras describen esta des-coordinación entre el calendario administrativo y el escolar:

Siempre se hace [comentar los resultados de las pruebas en el claustro], lo que ocurre es que llegan muy tarde. Llegaron, por ejemplo, la semana pasada y sí que hicimos una reunión y entonces [fue] la única reflexión que nos dio tiempo de hacer... pero después sí que cuando encaramos el trabajo con los grupos entonces hacemos una más específica. Pero al final es el tema de reflexionar sobre estas formas de trabajar en el aula, qué resultados nos han ayudado a obtener, hacemos esta reflexión (Docente 2, escuela pública 1)

Gráfico 3 – Uso de los datos de las pruebas de competencias básicas (docentes)



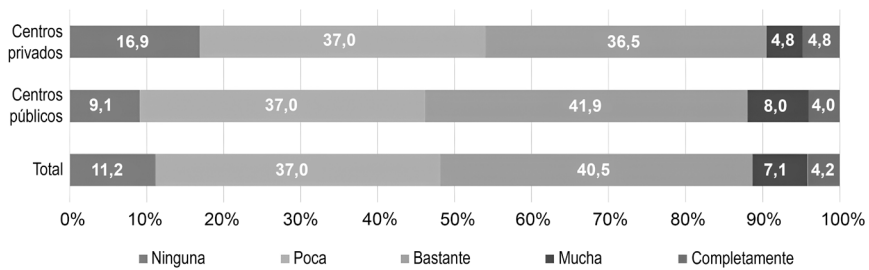
Fuente: Proyecto Reformed (Cataluña).

El limitado uso de los resultados de las pruebas de competencias básicas también obedece a que dichos datos no siempre se consideran indicadores fiables de calidad educativa (cuestión que abordamos más adelante) y a las dificultades de los centros para convertirlos en propuestas concretas. De hecho, poco más de la mitad del profesorado catalán declara tener al menos cierto grado de dificultad para transformar estos datos en medidas específicas de mejora (gráfico 4). Esta percepción varía según la titularidad: mientras el 54% de los docentes de centros concertados señalan poca o ninguna dificultad, en la pública este porcentaje desciende al 48%.

Asimismo, la complejidad social del centro influye en estas dificultades (gráfico 5): en los centros de composición alta, seis de cada diez docentes encuentran poca o ninguna barrera para traducir los datos en acciones; en los de composición media y baja, este porcentaje cae al 47% y al 40%, respectivamente. Esto puede deberse tanto a la mayor presión por mejorar resultados en entornos desfavorecidos, que expone

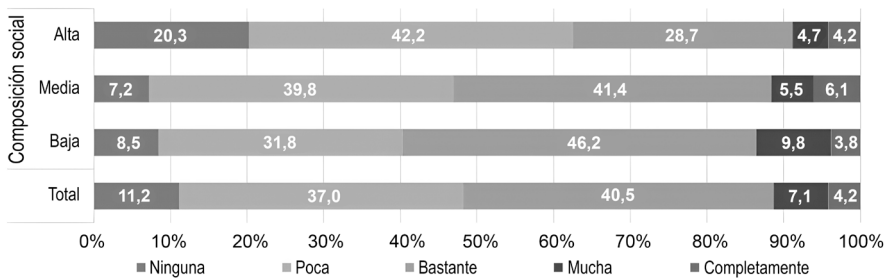
las limitaciones a la hora de incorporar los datos en planes de mejora, como a diferencias en prioridades y disponibilidad de tiempo para trabajar aspectos concretos de la enseñanza.

Gráfico 4 – Dificultad para transformar los datos provenientes de las pruebas de CB en medidas y acciones para mejorar la enseñanza, por titularidad (docentes)



Fuente: Proyecto Reformed (Cataluña).

Gráfico 5 – Dificultad para transformar los datos provenientes de las pruebas de CB en medidas y acciones para mejorar la enseñanza, por niveles de composición social (docentes)



Fuente: Proyecto Reformed (Cataluña).

El 40% del profesorado señala la falta de tiempo como el principal obstáculo para traducir los datos en acciones, cifra que desciende al 33% en centros de composición social aventajada y asciende al 42% en los de composición social media y desaventajada (gráfico 6). Casi la mitad

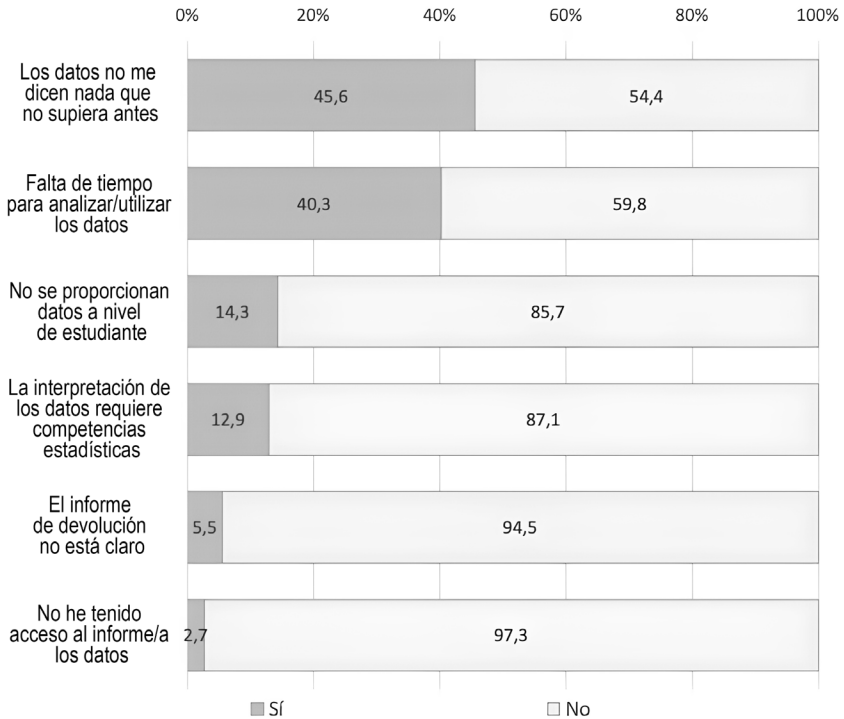
de los docentes añade que los resultados de las pruebas externas no les aportan información nueva. Esta percepción se refleja en las siguientes declaraciones de dos maestras, una de una escuela con población más desaventajada y una de una concertada:

P: ¿Le ayudan [los resultados de las pruebas] a identificar carencias de la escuela o algunos puntos donde se deba mejorar algo? R: Es que las carencias de esta escuela ya las sabemos por el tipo de escuela que somos y las intentamos mejorar cada día... (Docente 27, escuela pública 9).

Los resultados de las competencias es, más o menos, lo que sabíamos y la información que nos da no es, 'ostras, qué sorpresa'. Puede pasar, puede variar, pero no nos da ningún dato nuevo, la verdad es ésta, no nos sorprende. Es verdad que lo que hacemos es buscar unos indicadores de éxito para superar aquellos vacíos que vemos de estas pruebas (Docente 36, escuela concertada 13).

El 14 % del profesorado señala la falta de datos a nivel de estudiante y el 13 % apunta que la interpretación exige competencias estadísticas. Solo un 6 % encuentra poco claro el informe de devolución y un 3 % cita problemas de accesibilidad, lo que sugiere que la documentación del CSASE es, en general, clara.

Gráfico 6 – Factores que explican las dificultades
percibidas por transformar los datos en medidas y
acciones concretas (docentes)



Fuente: Proyecto Reformed (Cataluña).

Factores asociados con el uso de los datos

Los datos de las evaluaciones externas pueden emplearse con fines formativos y expresivos. Por ejemplo, para evaluar el trabajo docente, detectar estudiantes con necesidades de apoyo, planificar el desarrollo profesional o ajustar el currículo. También pueden utilizarse con propósitos sumativos e instrumentales, como informar a las familias, separar al alumnado por nivel, comparar rendimientos con otras escuelas o mejorar la reputación del centro. En este apartado, analizamos qué factores organizativos (estructura, recursos, liderazgo) e individuales (perfil y competen-

cias del profesorado) están asociados a estos distintos usos de los datos dentro de los centros educativos.

Uso de los datos de carácter instrumental

Nuestro análisis revela que la presión percibida para lograr buenos resultados en las pruebas de competencias básicas se asocia con un uso de los datos de corte competitivo: comparar rendimientos con otros centros y reforzar la reputación del centro (tabla 1). Esta presión también vincula los datos con la evaluación del trabajo docente y, en menor medida, con la agrupación de alumnado por niveles, lo que sugiere la existencia de una cultura escolar que instrumentaliza la información. Este patrón es especialmente marcado en la red concertada: tras ajustar por presión, cultura de evaluación y composición escolar, el profesorado de centros concertados duplica las probabilidades de emplear los datos para compararse con otras escuelas o mejorar su reputación, y los utiliza en mayor medida para informar a las familias.

En cambio, en la pública es más frecuente el uso de datos, de cualquier origen, no solo de CB, para agrupar a los estudiantes con miras a optimizar sus resultados. Además, la confianza en la validez de las pruebas se correlaciona de forma notable con estos usos instrumentales y competitivos. Por nivel educativo, el profesorado de secundaria recurre con más frecuencia a los datos para realzar la imagen del centro y evaluar la acción docente, mientras que en primaria se emplean con mayor intensidad para fundamentar decisiones de formación y desarrollo profesional.

Uso de los datos de carácter formativo

Nos centramos ahora en los factores que promueven o dificultan un uso formativo de los datos orientado a la mejora pedagógica. El principal factor identificado que obstaculiza su uso son las dificultades percibidas por el profesorado para transformar los datos en medidas y acciones concretas de mejora. En este sentido, el profesorado que tiene dificultades para traducir los datos en acciones de mejora, tiene casi la mitad de probabilidad de utilizar los datos para ajustar el currículum o tomar

decisiones sobre el desarrollo profesional. Además, cabe destacar que la dificultad percibida no está asociada a ninguna otra finalidad del uso de los datos. Sin embargo, como vemos en el gráfico 6, las dificultades no son de naturaleza técnica, sino que se deben tanto a la falta de utilidad percibida de los datos para iniciar procesos de mejora, como a las dificultades asociadas a la falta de tiempo y la carga de trabajo.

Por otra parte, el uso de los datos con fines formativos también está asociado con dinámicas de colaboración del profesorado en materia didáctica y pedagógica. Así, observamos que la frecuencia de estas dinámicas de colaboración está relacionada con la probabilidad de hacer un uso de los datos tanto para tomar decisiones en materia de formación como para informar a los demás profesores y profesoras sobre los resultados de las pruebas.

Tabla 1 – Presión por obtener buenos resultados y uso de los datos de las pruebas de competencia (Continúa)

	(1)	(2)	(3)	(4)
	Uso de los datos para evaluar el trabajo docente	Uso de los datos para identificar a estudiantes con necesidad de apoyo	Uso de los datos para tomar decisiones de desarrollo profesional	Uso de los datos para ajustar el currículum
Presión para obtener buenos resultados	1.46** (0.24)	0.98 (0.12)	1.02 (0.14)	1.20* (0.13)
Cultura favorable a las pruebas	1.09 (0.17)	1.15* (0.09)	0.98 (0.13)	0.94 (0.10)
Dificultad para transformar los datos	0.85 (0.25)	0.65 (0.22)	0.60* (0.18)	0.58** (0.16)
Edad	0.97*** (0.01)	0.99 (0.01)	0.99 (0.01)	1.03*** (0.01)
Género (ser mujer vs. Hombre)	0.70 (0.19)	0.64** (0.12)	0.74* (0.12)	1.29 (0.31)

Tabla 1 – Presión por obtener buenos resultados y uso de los datos de las pruebas de competencia (continúa)

Nivel educativo (docentes secundaria vs. Primaria)	1.79** (0.51)	0.92 (0.29)	0.47*** (0.12)	1.04 (0.20)
Titularidad (escuela pública vs. Privada)	0.97 (0.22)	1.22 (0.28)	0.89 (0.17)	1.11 (0.24)
Competencia performativa entre profesores	1.28*** (0.12)	1.06 (0.08)	1.07 (0.11)	0.98 (0.08)
Colaboración con los compañeros	0.93 (0.14)	0.83 (0.14)	1.51* (0.35)	0.74 (0.15)
Observaciones	652	652	655	655
Pseudo R 2	0.051	0.022	0.037	0.039

	(5)	(6)	(7)	(8)	(9)
	Uso de los datos para informar a las familias	Uso de los datos para agrupar a los estudiantes	Uso de los datos para comparar el rendimiento con otras escuelas	Uso de los datos para favorecer la reputación de la escuela	Uso de los datos para informar a los profesores
Presión para obtener buenos resultados	0.87 (0.09)	1.71* (0.54)	1.33** (0.19)	1.58** (0.33)	1.21 (0.16)
Cultura favorable a las pruebas	1.21* (0.14)	1.42* (0.27)	1.13 (0.10)	1.41*** (0.17)	1.09 (0.10)
Dificultad para transformar los datos	0.91 (0.21)	1.15 (0.64)	1.41 (0.34)	1.25 (0.35)	1.25 (0.35)
Edad	1.02*** (0.01)	1.04* (0.02)	0.99 (0.01)	0.97*** (0.01)	0.99* (0.01)
Género (ser mujer vs. Hombre)	1.26 (0.18)	1.25 (0.59)	0.99 (0.17)	0.81 (0.27)	0.88 (0.16)
Nivel educativo (docentes secundaria vs. Primaria)	0.81 (0.19)	1.55 (1.19)	0.89 (0.23)	2.54*** (0.56)	0.85 (0.22)

Tabla 1 – Presión por obtener buenos resultados y uso de los datos
de las pruebas de competencia (conclusión)

Titularidad (escuela pública vs. Privada)	0.42*** (0.10)	3.60** (2.27)	0.58* (0.18)	0.53* (0.19)	0.66 (0.20)
Competencia performativa entre profesores	1.09 (0.09)	1.16 (0.15)	1.19** (0.09)	1.16 (0.13)	0.95 (0.08)
Colaboración con los compañeros	1.09 (0.25)	0.90 (0.35)	0.87 (0.16)	0.94 (0.24)	1.33* (0.23)
Observaciones	655	627	655	655	655
Pseudo R 2	0.057	0.082	0.051	0.106	0.029

Nota: Los valores corresponden a los *odds ratio* de uso de los datos de las pruebas. Cada columna refleja un modelo de regresión logística con efectos fijos, en el que las distintas variables de uso de los datos son función de una batería de variables individuales y del centro educativo. Además de las variables mostradas, se incluyeron como controles: proporción de estudiantes extranjeros (composición social), proporción de estudiantes repetidores (desventaja educativa), carga de trabajo (horas), efecto fijo de la provincia de Barcelona y *missing dummies* para variables con valores perdidos (Cohen; Cohen, 1983; OCDE, 2007). Los coeficientes están exponentiados (*odds-ratio*); los errores estándar aparecen entre paréntesis.

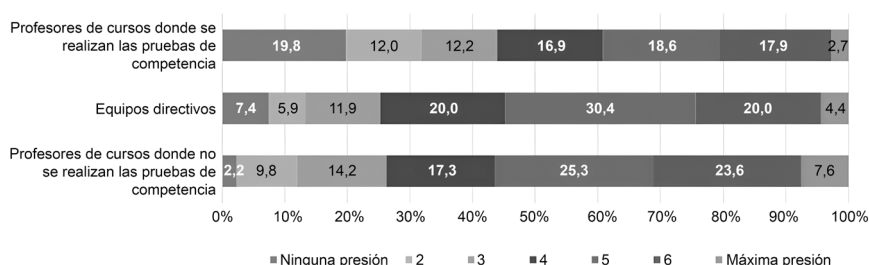
Presión por rendir en las pruebas de competencias básicas

Una vez descrito el uso limitado de los datos de evaluación externa, así como los usos formativos e instrumentales identificados y los factores que restringen su aprovechamiento pedagógico, este apartado aborda el segundo eje de resultados: la presión que experimentan los centros en un sistema formalmente *low-stakes*. Como se ha presentado, en el sistema de evaluación escolar catalán, las pruebas de competencias básicas no llevan aparejadas sanciones económicas ni la publicación de resultados escolares, y su influencia en los procesos de promoción docente es mínima. Se utilizan principalmente como un indicador de referencia en los planes de mejora escolar, y sirven como uno de los elementos para valorar la labor del equipo directivo, aunque siempre en el marco de un conjunto de indicadores mucho más amplio. Por ello, se considera que es un sistema *low-stakes*.

Sin embargo, que el sistema sea formalmente *low-stakes* no impide que docentes y directivos experimenten presiones derivadas de factores más sutiles: la reputación del centro ante las familias, la valoración interna del cuerpo de profesores/as, y los procesos de mejora escolar y evaluación directiva, en el marco de los cuales la inspección y la administración educativa hacen cada vez más hincapié en los resultados de CB para clasificar y exigir planes de mejora a los centros. Por lo tanto, aun sin consecuencias formales, la percepción de que ‘la imagen del centro’ o ‘el prestigio profesional’ depende de esos datos genera un ambiente de exigencia. El progresivo descenso en los resultados de las pruebas de CB, junto con el bajo rendimiento del sistema educativo catalán en evaluaciones internacionales, especialmente en comparación con la media europea y con países considerados de referencia, ha contribuido a reforzar esta tendencia (Quilabert *et al.*, 2024).

Los datos recogidos revelan que esta presión percibida es considerable. Según se presenta en el gráfico 7, aproximadamente el 40% del profesorado cuyos alumnos afrontan las pruebas sitúa su nivel de presión en las categorías más altas (5–7 en una escala de 1 a 7), mientras que únicamente un 20 % declara no percibir ninguna presión. Entre los equipos directivos, la tensión se acentúa: más de la mitad señala niveles elevados de presión, y sólo alrededor del 7% afirma no sentirse presionado. Este desajuste entre el carácter *low-stakes* formal y la presión real sugiere que los mecanismos informales de responsabilidad (reputación, presión de pares, creciente interés en los medios por los resultados de las CB, etc.) pueden ejercer un efecto performativo similar al de las sanciones directas, lo cual acaba repercutiendo en la cultura escolar y en las prácticas educativas que se llevan a cabo en los centros.

Gráfico 7 – Nivel de presión para obtener buenos resultados
en las pruebas (profesorado y equipos directivos)

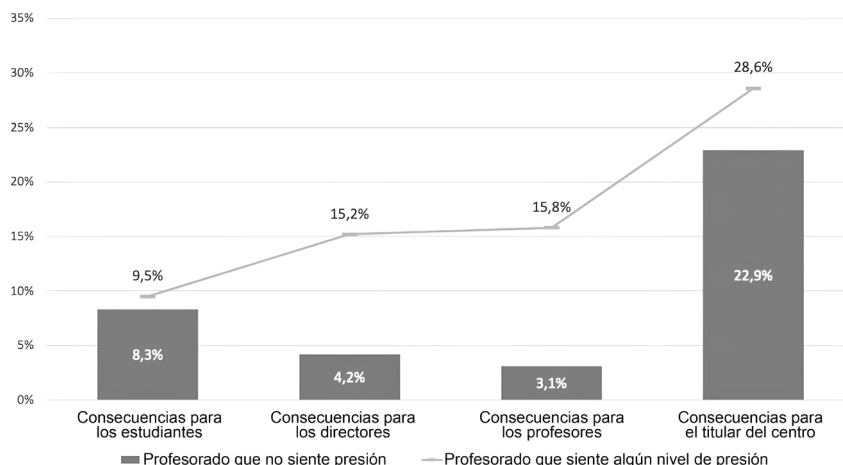


Fuente: Proyecto Reformed (Cataluña).

En el gráfico 7 también observamos que, de forma contraintuitiva, la presión por rendir no se concentra necesariamente en quienes trabajan en cursos evaluados mediante las pruebas de competencias básicas, sino que resulta más intensa entre quienes no participan directamente en ellas. Mientras que un 40% de los docentes de cursos evaluados reporta niveles altos o muy altos de presión, esta proporción asciende a casi el 60% entre quienes se desempeñan en cursos no evaluados. Igualmente llamativa es la diferencia en la ausencia total de presión: prácticamente ningún maestro de un curso no evaluado declara no sentir exigencia alguna, frente a dos de cada diez profesores de cursos evaluados que sí perciben un entorno libre de exigencias.

Este último grupo atribuye su sensación de ‘cero presión’ a la convicción de que no existen riesgos, ni reputacionales ni profesionales, asociados a los resultados: solo un 3 % cree que puedan derivarse consecuencias negativas para el profesorado, frente al 16% de quienes admiten al menos una presión mínima (Gráfico 8). Estos datos sugieren que, en contextos *low-stakes*, la presión puede desplazarse hacia docentes de cursos no evaluados, posiblemente por dinámicas internas de centro, expectativas implícitas, responsabilidades indirectas en la preparación del alumnado evaluado o simplemente falta de información sobre las consecuencias de las pruebas. No obstante, este patrón contraintuitivo plantea interrogantes que requieren una exploración más profunda para comprender sus causas y mecanismos.

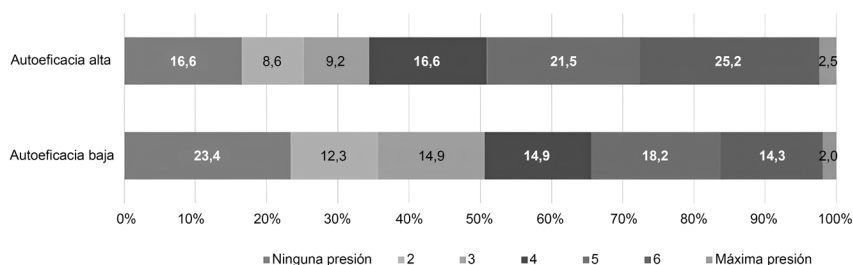
Gráfico 8 – Profesores que consideran que las pruebas de competencias tienen algún tipo de consecuencia (económicas, relativas al trabajo, reputacional, etc.), según nivel de presión percibida



Fuente: Proyecto Reformed (Cataluña).

La presión percibida también guarda relación con la autoeficacia del profesorado, esto es, la confianza en su capacidad para gestionar eficazmente el aula, mejorar el contenido de sus lecciones y mantener relaciones comunicativas positivas con el alumnado. Quienes creen firmemente que pueden impulsar cambios en su centro son, además, quienes más sienten la presión por los resultados, probablemente porque consideran que estos dependen en buena medida de su propio desempeño y no solo de factores externos. Así, cerca del 50% de los docentes con alto nivel de autoeficacia reporta sentir niveles elevados de presión, frente al 34% de aquellos que no se consideran capaces de gestionar la clase con eficacia (gráfico 9).

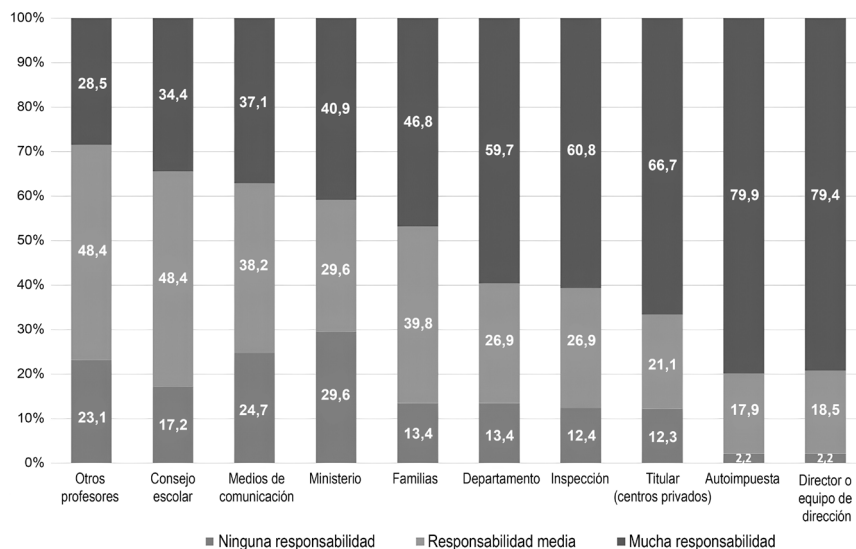
Gráfico 9 – Presión por obtener buenos resultados
y autoeficacia percibida por el profesorado



Fuente: Proyecto Reformed (Cataluña).

Una vez establecido el nivel de presión performativa, analizamos sus fuentes entre el profesorado que declara estar en las categorías más altas (5–7 del gráfico 9). Como muestra el Gráfico 10, tras la presión autoimpuesta, los principales responsables percibidos son los equipos directos y, en los centros privados, los titulares de la escuela. Así, el 79% de quienes sienten mucha presión atribuye parte de ella al equipo directivo, y el 67% a la titularidad del centro. Además, seis de cada diez docentes apuntan al Departamento de Educación y a la Inspección como fuentes significativas de esa misma presión para lograr buenos resultados en las pruebas de CB. Por lo general, los agentes a los que se les atribuye menor responsabilidad como fuentes de presión son el consejo escolar y los/as compañeros/as.

Gráfico 10 – Nivel de responsabilidad de los agentes por la presión por resultados percibida por el profesorado (docentes que perciben un alto nivel de presión)



Fuente: Proyecto Reformed (Cataluña).

Aunque el equipo directivo constituye la principal fuente de presión percibida, tanto los datos del cuestionario como las entrevistas indican que el profesorado no experimenta una presión explícita por parte de la dirección para orientar la enseñanza hacia los resultados. Más bien, esta exigencia se manifiesta como una expectativa compartida que emerge del clima de opinión existente entre el profesorado del centro y en el propio equipo directivo, pero que raramente se traduce en directrices concretas o indicaciones formales. Así lo plantean dos docentes de una escuela pública, entrevistadas conjuntamente:

R2: Yo pienso que partimos de un equipo directivo que, por ejemplo, ya sabe cómo trabajamos las dos tutoras que hay en sexto; por tanto, confía en nosotras, ni nos presionan ni nos dicen 'es muy importante'.

R1: Es que también nosotras sabemos lo que es importante, pero yo en ningún momento, ni tú tampoco, hemos sido presionadas. Ellos saben que sabemos que es importante; si suspendiéramos las pruebas, supongo que el equipo directivo diría ‘escucha, que estas pruebas son importantes, que estamos a sexto’, pero yo pienso que no es el caso.

R2: Yo pienso que hay dos personas que saben cómo somos y confían en nosotras; si no, no nos hubieran puesto a sexto, habrían puesto a otros maestros. Es lo que a mí me transmiten: que confían plenamente, saben cómo somos y cómo trabajamos, y que saben que haremos lo que podamos y que ya lo estamos haciendo, de hecho. Por tanto, este aspecto es muy importante: que un equipo directivo te transmita esto (Docentes 26 y 27, escuela pública 9).

El profesorado experimenta un alto nivel de presión autoimpuesta: el 80% de quienes sienten un alto nivel de exigencia identifica este tipo de presión. Sin embargo, la presión autoimpuesta no se da en el vacío. Podría reflejar, en gran medida, la internalización de demandas y expectativas externas. Como se presenta en la tabla 2, el 38% de la variabilidad de la presión autoimpuesta coincide con la presión percibida por parte del equipo directivo y, en los centros privados, un 20% con la presión del titular del centro, lo que sugiere que cuanto más presión ejerzan estos agentes, mayor es el grado en que los docentes “se exigen” a sí mismos.

Otras fuentes vinculadas a la presión autoimpuesta son las familias (27%), otros compañeros (23%), el consejo escolar (20%) y la inspección (17%). En conjunto, estos datos indican que la presión que los docentes sienten como propia se alimenta de una mezcla de expectativas profesionales internas y de las demandas que perciben en su entorno educativo, que va desde la dirección y la titularidad hasta el nivel de implicación de la comunidad escolar.

Tabla 2 – Relación entre la presión autoimpuesta y la presión que proviene de otras fuentes (docentes)

Otras fuentes de presión:	Presión autoimpuesta	
	Coeficiente de correlación	Variabilidad compartida
Director o equipo de dirección	0.615	38%
Familias	0.516	27%
Otros profesores	0.479	23%
Titular (centros privados)	0.452	20%
Consejo escolar	0.451	20%
Inspección	0.415	17%
Departamento	0.378	14%
Medios de comunicación	0.335	11%
Ministerio	0.298	8%

Nota: Todas las correlaciones son significativas al 1%.

Fuente: Proyecto Reformed (Cataluña).

Complementariamente, de las entrevistas se desprende que la presión autoimpuesta deriva de una combinación de autoexigencia, interés por el aprendizaje del alumnado y mantenimiento de cierta reputación profesional frente a los iguales. Una docente de una escuela de composición social media-baja lo plantea con estas palabras:

P: ¿De dónde dirías que viene esa presión que describes?

R: Bueno, de mí, ¿no? Porque yo soy exigente y quiero que los niños escriban bien y dices: 'vaya, si he estado trabajando los plurales, pues que me lo hagan bien', porque muchas veces lo has trabajado mucho y ese día no lo aplican, y entonces parece que no lo hayas trabajado, ¿sabes? [...]

P: ¿Dirías que es una sensación similar para el resto de docentes o piensas que hay quien no la siente?

R: Yo creo que sí, porque tengo conocidos que dedican tiempo a preparar las pruebas.

P: ¿Y en este centro?

R: Claro, es que el resto no, porque solo es el de 6.º... claro, quizá sea una sensación mía, ¿eh?, pero cuando salen los resultados siempre miran al maestro de sexto; pero claro, al resto no creo que les preocupe...

P: Tú vas haciendo un año 5.º y un año 6.º, ¿te sientes diferente un año respecto al otro?

R: Por eso te decía que no hay presión porque no te dicen nada, pero sabes que si salen mal, te mirarán a ti (Docente 45, escuela pública 15).

Presión de las evaluaciones externas: impacto en prácticas escolares y efectos colaterales

En este apartado examinamos cómo la presión derivada de las evaluaciones externas se relaciona con diversas prácticas pedagógicas: por un lado, con aquellas habitualmente señaladas como efectos colaterales de las políticas de rendición de cuentas (*teaching to the test* — TTT, reducción del currículo y agrupación del alumnado por niveles de desempeño), por otro, con el uso de datos con fines de mejora escolar. No obstante, mientras que las agrupaciones homogéneas de rendimiento preocupan por sus impactos negativos en equidad y cohesión social, el TTT solo se considera problemático cuando monopoliza el tiempo lectivo o se aplica de forma mecánica y sin *feedback* adecuado.

Organizamos este apartado en tres bloques. Primero, exploramos la asociación entre niveles de presión percibida y determinadas prácticas escolares. A continuación, profundizamos en cómo distintas fuentes de presión influyen en la práctica específica de *teaching to the test*. Por último, analizamos otros factores que, más allá de la presión percibida, también se vinculan con determinadas prácticas y efectos colaterales.

Fuentes de presión por los resultados y prácticas educativas: una exploración

El profesorado que siente mayor presión para obtener buenos resultados en las pruebas de CB tiende a practicar más *teaching to the test* y a alinear el currículo escolar con los contenidos de las pruebas (véase la tabla 3). Esta asociación es especialmente fuerte cuando la presión proviene de otros maestros, de las familias y, sobre todo, del equipo directivo (y, en los centros concertados, también del titular). Aunque menos intensa, la relación persiste cuando la fuente de presión es externa al centro (Departamento o Inspección) y cuando es autoimpuesta, lo que sugiere una cultura de responsabilidad compartida. Por el contrario, la vinculación entre presión percibida y uso de los datos de las pruebas de CB (por ejemplo, para diseñar planes de mejora) es más débil, y destaca únicamente un uso ligeramente mayor cuando la presión procede del equipo directivo (y del titular en la concertada).

Conviene señalar, además, que la simple conciencia de posibles consecuencias derivadas de los resultados no se asocia estrechamente con prácticas de TTT o reducción curricular, salvo cuando el profesorado percibe que el equipo directivo ejerce esa presión. Algunas entrevistas apuntan a que la incertidumbre sobre el alcance real de esas consecuencias alimenta la preparación intensiva para las pruebas en un sistema formalmente *low-stakes*, pero percibido como exigente. Esto se ilustra en la siguiente cita:

P: ¿Sientes presión por parte de las pruebas de competencias básicas?

R: Nosotras sí. Tenemos presión porque son unos números: si salen bajos, nos preguntamos qué consecuencias tendrá esto... (Docente 21, escuela pública 7).

Por último, a diferencia de lo que sucede en otros países, la presión atribuida a los medios de comunicación no muestra ninguna asociación con prácticas tácticas de optimización de resultados.

Tabla 3 – Correlación entre el grado de responsabilidad
atribuida a los agentes de la presión por resultados
y las prácticas escolares (docentes) (continúa)

		Ministerio	Departamento	Inspección	Titular (privados)	Director y equipo directivo	Consejo escolar	Compañeros/as	Familias	Autoimpuesta	Medios comunicación
Uso de los resultados de las pruebas	Ajustar el currículum	0.07	0.10	0.09	0.10	0.10	0.10	0.09	0.08	0.11	0.05
	Informar a los padres sobre el rendimiento escolar	-0.03	-0.01	-0.03	0.26	0.12	0.05	0.09	0.11	0.14	0.05
	Agrupar a los estudiantes	0.06	0.07	0.06	0.04	0.10	0.08	0.14	0.11	0.07	0.08
	Comparar el rendimiento con el de otras escuelas	0.01	0.02	0.02	0.25	0.16	0.04	0.08	0.04	0.09	0.05
	Favorecer la reputación de la escuela	-0.07	-0.07	-0.04	0.19	0.21	0.08	0.11	0.09	0.08	-0.02
	Informar al profesorado	-0.01	0.00	0.01	0.14	0.08	0.00	0.05	0.03	0.06	0.01
Reducción / alineación del currículum	Las pruebas de CB llevan a una redistribución de los recursos	0.19	0.20	0.20	0.32	0.33	0.26	0.28	0.24	0.20	0.11
	Los estándares de aprendizaje influyen en el enfoque pedagógico	0.14	0.21	0.21	0.33	0.33	0.21	0.24	0.23	0.22	0.11
	Se consideran las pruebas de CB por las decisiones curriculares	0.12	0.19	0.23	0.31	0.38	0.27	0.29	0.26	0.28	0.10

Tabla 3 – Correlación entre el grado de responsabilidad atribuida a los agentes de la presión por resultados y las prácticas escolares (docentes) (conclusión)

Teaching to the test	Frecuencia de actividades para preparar a los estudiantes	0.05	0.11	0.16	0.18	0.16	0.09	0.06	0.13	0.26	0.04
	Actividades para preparar a los estudiantes durante todo el año	0.13	0.16	0.17	0.11	0.20	0.09	0.10	0.06	0.22	0.14
	Actividades para preparar a los estudiantes el mes anterior de la prueba	0.09	0.14	0.15	0.25	0.23	0.10	0.09	0.12	0.31	0.07
Conocimiento de las consecuencias de las pruebas	Consecuencias de los resultados de las pruebas para el director	0.08	0.15	0.17	0.15	0.20	0.12	0.15	0.13	0.17	0.15
	Consecuencias de los resultados de las pruebas para el profesorado	0.02	0.08	0.12	0.24	0.19	0.06	0.10	0.10	0.18	0.07
	Consecuencias de los resultados de las pruebas para el titular de la escuela	-0.01	0.07	0.13	0.11	0.15	0.06	0.06	0.10	0.16	0.07

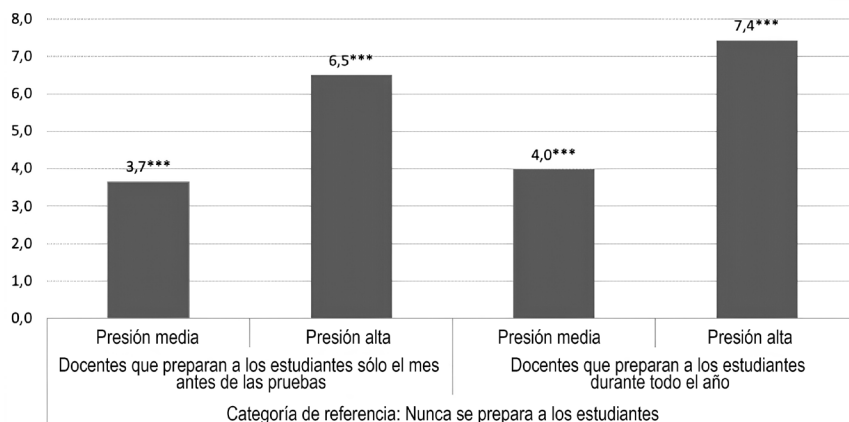
Fuente: Proyecto Reformed (Cataluña).

Preparación de los estudiantes para las pruebas según niveles de presión percibida

En este subapartado, relacionamos la presión percibida para obtener buenos resultados en las pruebas de competencias básicas con la práctica del *teaching to the test*, es decir, la preparación específica del alumnado para afrontar con éxito los tests estandarizados. Para ello, comparamos la probabilidad (riesgo) de que el profesorado que siente presión, en

comparación con el que no la percibe, haya dedicado tiempo a preparar a sus estudiantes durante el mes previo a las pruebas o a lo largo de todo el curso. Los resultados de los modelos, ilustrados en el Gráfico 11, apuntan que quienes reportan niveles medios o altos de presión performativa adoptan con mayor frecuencia prácticas de TTT. Concretamente, un docente que percibe alta presión tiene hasta 7,4 veces más probabilidad de realizar TTT durante todo el año que uno que no siente presión. Esta relación se mantiene, aunque en menor medida (odds ratio = 6,5), cuando se analiza únicamente la preparación en el mes anterior a las pruebas.

Gráfico 11 – Teaching to the test y presión para obtener buenos resultados en las pruebas de competencias básicas (odds ratio)



Modelos de regresión logística multinomial. Variable respuesta: preparar a los estudiantes para las pruebas de competencias básicas (nunca / mes antes / durante todo el año). Los valores ilustrados corresponden a los odds ratio de preparar a los estudiantes para las pruebas, según la frecuencia y la presión percibida para obtener buenos resultados. Variables de control a nivel individual: experiencia, tipo de contrato, afiliación sindical, nivel educativo, género, satisfacción con el trabajo y con la escuela, enfoque tradicional / innovador, preferencia por metodologías metacognitivas. Control a nivel de centro: efecto fijo de centro educativo. Signif. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Pseudo R^2 : 0.32.

Fuente: Proyecto Reformed (Cataluña).

Además de la presión, exploramos si la posición de la escuela en su mercado educativo local (MEL; determinado por su reputación, rendimiento

medio y nivel de demanda) también influye en la probabilidad de practicar *teaching to the test*. El gráfico 12 revela que, junto a la presión, la posición en el MEL se asocia con prácticas de TTT: los centros con mala posición que además perciben alta presión muestran una mayor predisposición a preparar a sus estudiantes que aquellos con igual posición pero menor presión. De igual modo, entre los centros bien situados en el MEL, la frecuencia de TTT aumenta o disminuye según el nivel de presión percibida, lo cual confirma que ambos factores operan de forma conjunta en la adopción de prácticas orientadas a optimizar los resultados de las pruebas.

Gráfico 12 – *Teaching to the test*, según niveles de presión y posición en el MEL (continúa)

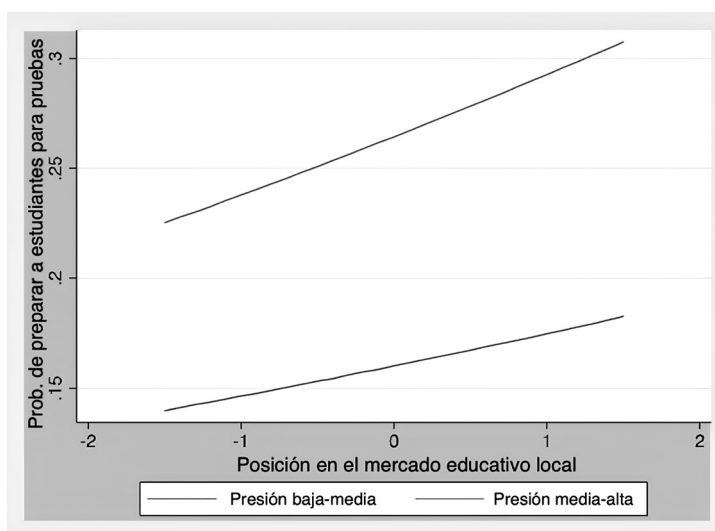
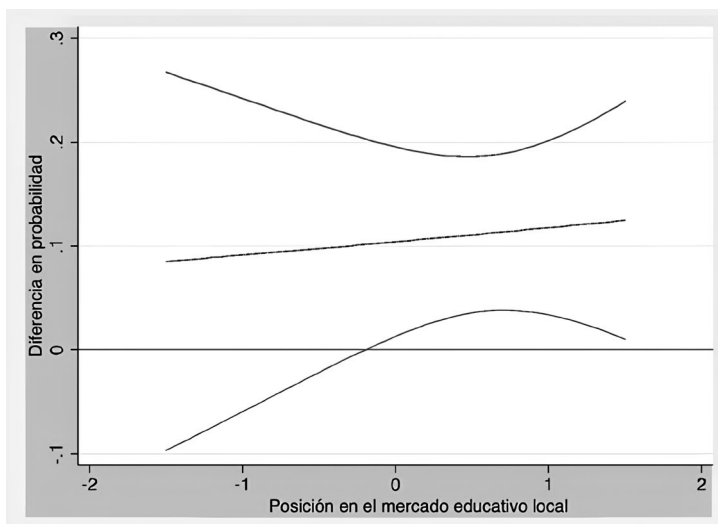


Gráfico 12 – *Teaching to the test*, según niveles
de presión y posición en el MEL (conclusión)



Nota: El gráfico muestra la interacción entre la posición en el MEL (variable continua) y el nivel de presión (variable categórica) en un modelo de regresión logística donde la probabilidad de realizar *teaching to the test* (0/1) depende de una batería de variables individuales y de centro: género, media jornada, afiliación sindical, edad, nivel educativo, titularidad, autonomía de gestión, índice de autonomía educativa, satisfacción con el trabajo, satisfacción con la escuela e índice de cultura escolar favorable a las pruebas. Las variables con valores perdidos se imputaron mediante missing dummies (Cohen; Cohen, 1983; OCDE, 2007). Signif. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fuente: Proyecto Reformed (Cataluña).

Cabe matizar que el *teaching to the test* no se percibe únicamente como una estrategia orientada a inflar los resultados. En algunos centros, estas pruebas se valoran positivamente por su potencial educativo: familiarizan al alumnado con formatos estandarizados, facilitan el entrenamiento en habilidades como la comprensión auditiva y ayudan a gestionar el tiempo de examen, especialmente en escuelas donde los exámenes escritos no son habituales. Desde esta perspectiva, la preparación para las pruebas puede entenderse como un modo de promover el bienestar del estudiante y reducir su ansiedad, más que como una práctica instrumental. En las entrevistas, varios docentes subrayan este enfoque formativo:

P: ¿Qué haces para intentar mejorar los resultados de las competencias básicas?

R: En sexto utilizamos algunos ejercicios de años anteriores: no hemos hecho muchas, pero las que sí practicamos ayudan por su tipología de actividades. Así evitamos que les sorprendan 'los listenings' o el tiempo limitado; en clase, a veces de manera inconsciente, pautamos las tareas, explicamos paso a paso, sin interrupciones, y luego hacemos prácticas en el tiempo estipulado (Docente 33, escuela pública 11).

Respecto a la carga de trabajo, corregimos y preparamos más a fondo: ya no es solo ensayar, sino explicar cada parte y calcular el tiempo disponible. Dedicamos muchas horas y sesiones a estas competencias porque los alumnos no suelen enfrentarse a un dossier tan extenso, de muchas páginas (Docente 30, escuela pública 12).

Rendición de cuentas y respuestas pedagógicas colaterales: otros factores explicativos

Más allá de la presión, otros factores inciden en prácticas como el *teaching to the test*, la reducción del currículo y la agrupación por niveles. En primer lugar, como muestra el modelo de la tabla 4, una cultura escolar que valora positivamente las pruebas de competencia se vincula con la redistribución de recursos hacia las materias evaluadas, la reducción curricular, el alineamiento con estándares de aprendizaje y la agrupación homogénea del alumnado. En segundo lugar, la presión por obtener buenos resultados se asocia con todas estas respuestas, excepto la agrupación. Las entrevistas confirmaron que, en escuelas donde se confía en la validez de las pruebas, estas no se usan solo de forma instrumental sino también como herramientas pedagógicas legítimas:

Lo que sí decimos es que realmente hay algunas pruebas que están bien hechas y entonces, si

esta forma de trabajar consideramos que es buena, pues vayamos por ahí, independientemente de que los resultados sean mejores o peores. Por ejemplo, la prueba de expresión escrita, lo que evalúa las rúbricas que tienen y la manera de poder desarrollarlo a nosotros nos parece muy bien y entonces nosotros lo hemos aplicado (Directora 14, escuela pública 14).

En tercer lugar, la composición social se asocia con una mayor intensidad de *teaching to the test*. El profesorado de estos centros recurre a simulacros regulares tanto para compensar las condiciones de partida del alumnado como para demostrar que, pese a las dificultades sociales, pueden ofrecer una educación de calidad.

P: ¿Hacéis ensayos de las pruebas?

R2: Sí, les llaman simulacros; vamos muy en serio.

R1: Hacen bastantes.

R2: E incluso utilizamos modelos de otros años, que se realizaron pruebas en cursos anteriores. De hecho, a partir del próximo año vuelve a cambiar...

R1: Sí, pero practicamos esas pruebas con antelación; no se las damos sin avisar, que sería un tema complicado de sexto. Les hablamos de ellas, les explicamos el formato.

P: Entiendo que tomáis elementos característicos de las competencias básicas e intentáis irlos introduciendo poco a poco en las clases, para que los alumnos se familiaricen.

R1: Sí, y después hacemos simulacros tal cual: tomamos una prueba antigua, la realizamos y revisamos los resultados (Directora 4 y Docente 12, escuela pública 4).

Tenemos descargadas pruebas de años anteriores, clasificadas en carpetas, y van practicando.

Una vez a la semana hacen un simulacro (Docente 26, escuela pública 9).

En cuarto lugar, un planteamiento pedagógico más centrado en el docente se asocia con una mayor frecuencia de preparación de los alumnos para las pruebas de CB y con la adaptación del currículo a sus contenidos. De manera más moderada, los enfoques tradicionales también muestran vínculos positivos con el alineamiento a los estándares de aprendizaje y la formación de agrupamientos homogéneos. Por ejemplo, en una escuela con un planteamiento educativo tradicional predominante, la directora y una docente consideran que los resultados de la evaluación externa son un indicador anual fundamental para la toma de decisiones en el centro:

P: ¿La prueba de competencias básicas es un instrumento que os sirve también a vosotros?

R1: Sí. Incluso ahora, cuando hacemos la memoria de final de curso, hay indicadores que nos sirven para entender cómo ha ido la expresión escrita de los niños; sacamos esta información de las pruebas de competencias.

R2: Lo que hacemos es comparar las notas que ponemos en la escuela, por ejemplo de expresión escrita, con la nota que obtienen los alumnos de sexto en expresión escrita, y luego comparamos: no es que estemos evaluando super alto, de qué van super bien y después llegan a la prueba y les sale muy bajo. En otras pruebas más estrictas ponían notas bajas, pero con las competencias dices escucha, pues lo hacen bastante bien; esto también es un buen análisis.

R1: La utilizamos y la contabilizamos como resultado.

R2: Y en los institutos también lo piden y lo utilizan de referencia (Directora 4 y Docente 12, escuela pública 4).

Finalmente, en quinto lugar, según la titularidad, las escuelas públicas tienden a concentrar recursos en las áreas evaluadas por las pruebas de CB, mientras que las concertadas responden a la presión por resultados intensificando la preparación específica para las pruebas. En las entrevistas en centros concertados, el profesorado describe cómo incorporan ejercicios estándar dentro del programa habitual y dedican tiempo a ensayos de exámenes anteriores:

P: ¿Hacéis alguna especie de ensayo de pruebas de años anteriores o dedicáis más horas, separáis en grupos o algo así?

R: Sí, te aseguro que sí. Ahora hablo de matemáticas: el programa que seguimos incluye día sí y día también ejercicios que podrían pertenecer a cualquier prueba de competencias básicas, pero lo preparamos con pruebas de años anteriores (Docente 35, escuela concertada 12).

Tabla 4 – Autonomía, niveles de presión y respuestas no esperadas (continúa)

	(1)	(2)	(3)	(4)	(5)
	Teaching to the test: actividades para preparar a los estudiantes (último mes antes de las pruebas)	Redistribución de los recursos en favor de las áreas evaluadas	El contenido de las pruebas se tiene en cuenta para tomar decisiones curriculares	Los estándares de aprendizaje influyen el enfoque pedagógico del centro	Agrupar a los estudiantes de forma homogénea
Autonomía educativa (currículo, métodos, material y formación)	0.10*	-0.08	-0.04	-0.04	-0.05
	(0.06)	(0.06)	(0.05)	(0.04)	(0.07)
Autonomía de gestión (salarios, selección docentes y directores, presupuestos)	-0.06	-0.01	-0.02	0.02	0.09
	(0.06)	(0.05)	(0.05)	(0.04)	(0.06)

Tabla 4 – Autonomía, niveles de presión y
respuestas no esperadas (conclusión)

Posición en el MEL	0.06	0.11*	0.10	0.07	-0.06
	(0.08)	(0.06)	(0.07)	(0.05)	(0.07)
Planteamiento pedagógico tradicional	0.18***	-0.01	0.08**	0.06	0.01
	(0.07)	(0.05)	(0.04)	(0.06)	(0.06)
Escuela pública (ref. privada)	-0.27**	0.28**	0.10	0.07	0.12
	(0.12)	(0.11)	(0.13)	(0.10)	(0.15)
Proporción de estudiantes extranjeros	1.29***	0.25	0.02	-0.22	-0.62
	(0.43)	(0.24)	(0.22)	(0.16)	(0.49)
Cultura escolar favorable a las pruebas	0.03	0.21***	0.09**	0.10**	0.17***
	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)
Presión para obtener buenos resultados	0.25***	0.29***	0.36***	0.20***	0.10
	(0.08)	(0.06)	(0.05)	(0.05)	(0.07)
Índice de autoeficacia	0.06	0.06**	0.03	0.03	0.07**
	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)
Observaciones	426	622	628	624	636
R2	0.111	0.129	0.169	0.079	0.048

Nota: Cada columna corresponde a un modelo OLS con efectos fijos en el que las diferentes prácticas o respuestas escolares son función de una batería de variables individuales y del centro educativo. Además de las variables mostradas, se incluyeron como controles: tipo de contrato laboral, afiliación sindical, nivel educativo, edad y efecto fijo de la provincia de Barcelona. Las variables con valores perdidos se imputaron mediante *missing dummy variables* (Cohen; Cohen, 1983; OCDE, 2007). Los errores estándar aparecen entre paréntesis. Signif. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. En la tabla, los niveles de significatividad se indican con colores claros para * $p < 0.10$ y con colores oscuros para ** $p < 0.05$ y *** $p < 0.01$.

Fuente: Proyecto Reformed (Cataluña).

Conclusiones

En este estudio hemos analizado cómo los centros educativos implementan, perciben y resignifican las políticas de evaluación externa y rendición de cuentas en el contexto de un sistema formalmente de bajas consecuencias (*low-stakes*) que, al mismo tiempo, sitúa la autonomía escolar como pieza central de la gobernanza educativa. A partir de un

enfoque mixto —que combina cuestionarios, entrevistas y análisis documental— mostramos que el carácter *low-stakes* de la política no garantiza, por sí solo, una experiencia de baja presión en los centros. De hecho, una proporción significativa de docentes y equipos directivos declara sentir una fuerte presión por obtener buenos resultados en las pruebas de competencias básicas, aunque esta presión no derive de incentivos o sanciones formales, sino de mecanismos informales vinculados a la reputación institucional, las expectativas de las familias, la autoexigencia profesional y el clima de opinión dentro de las propias comunidades docentes.

Este hallazgo cuestiona el supuesto de que sistemas *low-stakes* conducen automáticamente a una apropiación formativa de la evaluación. Nuestros resultados muestran que la presión performativa se asocia con prácticas más instrumentales —como el *teaching to the test* y el estrechamiento curricular—, que desdibujan el propósito diagnóstico y formativo de la evaluación y tienden a limitar la enseñanza a los contenidos evaluados. Estos efectos aparecen con mayor intensidad en centros que concentran alumnado vulnerable y se encuentran en posiciones más desfavorecidas en el mercado educativo local, lo que evidencia que la competencia entre escuelas puede erosionar el carácter *low-stakes* del sistema y reforzar desigualdades ya existentes.

Frente a visiones que reducen los efectos no deseados a la presencia de incentivos materiales, nuestros hallazgos subrayan que, incluso en ausencia de sanciones explícitas, los mecanismos simbólicos de responsabilización y rendición de cuentas pueden producir dinámicas similares a las observadas en contextos *high-stakes*. Además, identificamos que la presión percibida no se concentra únicamente en quienes enseñan en cursos evaluados, sino que también se desplaza hacia docentes de cursos no sometidos a pruebas, un patrón contraintuitivo que sugiere la existencia de expectativas implícitas y responsabilidades compartidas dentro de los centros.

A su vez, el estudio destaca el papel clave de la cultura escolar y de la legitimidad de las pruebas en la configuración de los usos de los datos. Cuando el profesorado percibe las evaluaciones externas como válidas, aumenta la probabilidad de que los datos se utilicen de manera formativa.

Sin embargo, la utilidad real de los datos suele verse limitada por la falta de tiempo, recursos y espacios institucionales para traducir los resultados en ajustes curriculares y pedagógicos. En entornos con mayor complejidad social esta brecha es más aguda, lo que revela un sesgo estructural en la capacidad para aprovechar el potencial de mejora asociado a la evaluación externa.

Asimismo, la implementación efectiva de estas políticas está fuertemente mediada por actores intermedios. La inspección educativa, en particular, desempeña un rol decisivo: cuando ejerce un acompañamiento pedagógico cercano, los centros reportan más reflexión y uso significativo de los datos; cuando la relación se reduce a supervisión burocrática, la evaluación tiende a ser percibida como un mecanismo de control y no como un recurso para la mejora.

En conjunto, estos resultados revelan que el impacto de las políticas de autonomía, evaluación y RdC depende menos de su diseño formal y más de su apropiación por parte de los actores escolares, así como de las condiciones institucionales que habilitan o constriñen su implementación. Incluso cuando se conciben explícitamente como herramientas de mejora, pueden generar efectos colaterales y resistencias si no se acompañan de una infraestructura sólida de apoyo, una comunicación clara sobre sus objetivos y una estrategia sostenida de desarrollo de capacidades docentes.

Desde una perspectiva de política educativa, este estudio pone de relieve la necesidad de reforzar el carácter formativo del sistema de evaluación en Cataluña. Ello implica avanzar hacia un marco normativo coherente, estable y comprensible; garantizar tiempos, recursos y apoyos adecuados para el análisis pedagógico de los datos; y consolidar estructuras intermedias capaces de articular la autonomía con una responsabilidad compartida orientada al aprendizaje de todo el alumnado. No se trata de fomentar la competencia entre centros ni de poner el foco en la rendición de cuentas individual, sino de promover una cultura de responsabilidad profesional colectiva que reconozca las desigualdades de partida y priorice el valor pedagógico de la evaluación sobre su función clasificatoria.

En definitiva, el caso catalán muestra que un sistema *low-stakes* puede operar de facto como un sistema de alta presión ambiental si no se acompaña de las condiciones necesarias para que los datos sean útiles, legítimos y transformadores. La evaluación externa, entendida como herramienta de conocimiento y justicia educativa, solo cumplirá su potencial cuando vaya unida a una apuesta clara por el apoyo institucional, la equidad y la mejora escolar en todas las escuelas, especialmente en aquellas que más lo necesitan.

Referências

ALLEN, R.; VIGNOLES, A. What should an index of school segregation measure? **Oxford Review of Education**, [S. l.], v. 33, n. 5, p. 643–668, 2007. DOI: <https://doi.org/10.1080/03054980701366306>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/03054980701366306>. Acesso em: 5 jul. 2025.

BALL, S. J.; MAGUIRE, M.; BRAUN, A. **How schools do policy**: Policy enactments in secondary schools. Abingdon: Routledge. 2011.

BARBANA, S.; DUMAY, X.; DUPRIEZ, V. Local implementation of accountability instruments in the French-speaking community of Belgium. **European Educational Research Journal**, [S. l.], v. 19, n. 2, p. 94–108. 2020. DOI: <https://doi.org/10.1177/1474904119850964>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1474904119850964>. Acesso em: 12 jul. 2025

BREAKSPEAR, S. **The policy impact of PISA**: An exploration of the normative effects of international benchmarking in school system performance. Paris: OECD Education Working Papers, 2012. DOI: <https://doi.org/10.1787/5k9fdfqffr28-en>. Disponível em: https://www.oecd.org/en/publications/the-policy-impact-of-pisa_5k9fdfqffr28-en.html. Acesso em: 3 jul. 2025

CLARKE, M.; LUNA-BAZALDUA, D. **Primer on large-scale assessments of educational achievement**. Washington, D.C.: World Bank. 2021. DOI: <https://doi.org/10.1596/978-1-4648-1659-8>. Disponível em: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/827991619500628075>. Acesso em: 14 de julho de 2025

COHEN, J.; COHEN, P. **Applied multiple regression/correlation analysis for the behavioural sciences**. Mahwah: Erlbaum, 1983.

COMISIÓN EUROPEA. **Education and training monitor 2018**. Luxemburgo: Publications Office of the European Union, 2018.

DATNOW, A.; PARK, V.; WOHLSTETTER, P. **Achieving with data**: How high-performing school systems use data to improve instruction for elementary students. Los Angeles: Center on Educational Governance; USC Rossier School of Education, 2007.

DE GRAUWE, A. Improving the quality of education through school-based management: Learning from international experiences. **International Review of Education**, [S. l.], v. 51, n. 4, p. 269–287, 2005. DOI: <https://doi.org/10.1007/s11159-005-7733-1>. Disponível em: <https://link.springer.com/article/10.1007/s11159-005-7733-1>. Acesso em: 2 jul. 2025

DEE, T. S.; JACOB, B. The impact of No Child Left Behind on student achievement. **Journal of Policy Analysis and Management**, [S. l.], v. 30, n. 3, p. 418–446, 2011. <https://doi.org/10.1002/pam.20586>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/pam.20586>. Acesso em: 9 jul. 2025.

FARRE, R. **Crònica de la Llei d'educació de Catalunya**. Catalunha: Fundació Bofill, 2009.

FERRER-ESTEBAN, G. **Autonomia de centre, segregació escolar i mercats en educació**: Una anàlisi internacional dels efectes de l'autonomia en l'equitat i l'eficàcia dels sistemes educatius. 2015. 269 f. Tese (Doutorado em Sociologia) – Faculdade de Sociologia e Políticas Públicas, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 2015.

GREANY, T. Innovation is possible, it's just not easy: Improvement, innovation and legitimacy in England's autonomous and accountable school system. **Educational Management Administration & Leadership**, [S. l.], v. 46, n. 1, p. 65–85, 2018. DOI: <https://doi.org/10.1177/1741143216659297>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1741143216659297>. Acesso em: 11 jul. 2025.

HAMILTON, L. S.; STECHER, B. M.; KLEIN, S. P. **Making sense of test-based accountability in education**. Santa Monica: Rand Corporation, 2002.

JERRIM, J.; SIMS, S. When is high workload bad for teacher wellbeing? Accounting for the non-linear contribution of specific teaching tasks. **Teaching and Teacher Education**, [S. l.], v. 105, 2021. DOI: <https://doi.org/10.1016/j.tate.2021.103395>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0742051X21001190?via%3Dihub>. Acesso em: 4 jul. 2025.

LEVATINO, A.; PARCERISA, L.; VERGER, A. Understanding the stakes: The influence of accountability policy options on teachers' responses. **Educational Policy**, [S. l.], v. 38, n. 1, p. 31–60. 2024. DOI: <https://doi.org/10.1177/08959048221142048>. Disponível em: <https://journals.sagepub.com/doi/10.1177/08959048221142048>. Acesso em: 14 jul. 2025

LEVINSON, B. A.; SUTTON, M.; WINSTEAD, T. Education policy as a practice of power: Theoretical tools, ethnographic methods, democratic options. **Educational Policy**, [S. l.], v. 23, n. 6, p. 767–795, 2009. DOI: <https://doi.org/10.1177/0895904808320676>. Disponível em: <https://journals.sagepub.com/doi/10.1177/0895904808320676>. Acesso em: 13 jul. 2025.

MALEN, B. Revisiting policy implementation as a political phenomenon. In: HONIG, M. I. (ed.). **New directions in education policy implementation: Confronting complexity**. Albany: State University of New York Press, 2006. p. 83–104.

MAROY, C.; VAN ZANTEN, A. Regulation and competition among schools in six European localities. **Sociologie du travail**, [S. l.], v. 51, p. 67–79, 2009. DOI: <https://doi.org/10.1016/j.socotra.2009.01.005>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0038029609000065?via%3Dihub>. Acesso em: 7 jul. 2025.

MAROY, C.; VOISIN, A. **Think piece on accountability**. Paris: UNESCO, 2017.

MITTMAN, J.; JENNINGS, J. L. Accountability, achievement, and inequality in American public schools: A review of the literature. In: SCHNEIDER, B. (ed.). **Handbook of the sociology of education in the 21st century**. Cham: Springer, 2018. p. 475–492.

NICHOLS, S. L.; HARRIS, L. R. Accountability assessment's effects on teachers and schools. In: BROWN, G. T. L.; HARRIS, L. R. (eds.). **Handbook of human and social conditions in assessment**. Abingdon: Routledge, 2016. p. 40–56.

OECD. **PISA 2006: Science competencies for tomorrow's world. Analysis** (vol. 1). Paris: OECD Publishing, 2007.

OECD. **PISA in focus**. n° 9. School autonomy and accountability: Are they related to student performance? Paris: OECD Publishing, 2011.

PAGES, M.; FERRER-ESTEBAN, G.; VERGER, A.; PRIETO, M. At the crossroad of performativity and the market: Schools' logics of action under a hybrid accountability regime. **Compare: A Journal of Comparative and International Education**, [S. l.], p. 1–21, 2023. DOI: <https://doi.org/10.1080/03057925.2023.2175316>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/03057925.2023.2175316>. Acesso em: 10 jul. 2025.

QUILABERT, E.; VERGER, A.; MOSCHETTI, M. C.; FERRER-ESTEBAN, G.; PAGÈS, M. The obstacle race to educational improvement: governance, policies, and practices in disadvantaged schools. **British Journal of Sociology of Education**, [S. l.], v. 45, n. 6, p. 934–956, 2024. DOI: <https://doi.org/10.1080/01425692.2024.2376596>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01425692.2024.2376596>. Acesso em: 6 jul. 2025.

SPILLANE, J. P.; DIAMOND, J. B.; BURCH, P.; HALLETT, T.; JITA, L.; ZOLTNERS, J. Managing in the middle: School leaders and the enactment of accountability policy. **Educational Policy**, [S. l.], v. 16, n. 5, p. 731–762, 2002. DOI: <https://doi.org/10.1177/089590402237311>. Disponível em: <https://journals.sagepub.com/doi/10.1177/089590402237311>. Acesso em: 1 jul. 2025.

TELTEMANN, J.; JUDE, N. Assessments and accountability in secondary education: International trends. **Research in Comparative and International Education**, [S. l.], v. 14, n. 2, p. 249–271, 2019. DOI: <https://doi.org/10.1177/1745499919846174>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1745499919846174>. Acesso em: 8 jul. 2025.

TERHART, E. Teacher resistance against school reform: Reflecting an inconvenient truth. **School Leadership & Management**, [S. l.], v. 33, n. 5, p. 486–500, 2013. DOI: <https://doi.org/10.1080/13632434.2013.793494>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/13632434.2013.793494>. Acesso em: 15 jul. 2025.

THIEL, C.; SCHWEIZER, S.; BELLMANN, J. Rethinking side effects of accountability in education: Insights from a multiple methods study in four German school systems. **Education Policy Analysis Archives**, [S. l.], v. 25, n. 93, 2017. DOI: <https://doi.org/10.14507/epaa.25.2662>. Disponível em: <https://epaa.asu.edu/index.php/epaa/article/view/2662>. Acesso em: 2 jul. 2025.

TORRES, R. Does test-based school accountability have an impact on student achievement and equity in education? A panel approach using PISA. **OECD Education Working Papers**, 250. OECD Publishing. 2021. DOI: <https://doi.org/10.1787/0798600f-en>. Disponível em: https://www.oecd.org/en/publications/does-test-based-school-accountability-have-an-impact-on-student-achievement-and-equity-in-education_0798600f-en.html. Acesso em: 5 jul. 2025.

UNESCO. **The promise of large-scale learning assessments**: Acknowledging limits to unlock opportunities. Paris: UNESCO, 2019. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000369947>. Acesso em: 14 jul. 2025.

VERGER, A.; FONTDEVILA, C.; PARCERISA, L. Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. **Discourse: Studies in the Cultural Politics of Education**, [S. l.], v. 40, n. 2, p. 248–270, 2019. DOI: <https://doi.org/10.1080/01596306.2019.1569882>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01596306.2019.1569882>. Acesso em: 9 de julho de 2025.

VOGLER, K. E. Comparing the impact of accountability examinations on Mississippi and Tennessee social studies teachers' instructional practices. **Educational Assessment**, [S. l.], v. 13, n. 1, p. 1–32. 2008. DOI: <https://doi.org/10.1080/10627190801968158>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/10627190801968158>. Acesso em: 7 jul. 2025.

WERLER, T.; FÆREVAAG, M. K. National testing data in Norwegian classrooms: A tool to improve pupil performance? **Nordic Journal of Studies in Educational Policy**, [S. l.], v. 3, n. 1, p. 67–81, 2017. DOI: <https://doi.org/10.1080/20020317.2017.1320188>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/20020317.2017.1320188>. Acesso em: 3 jul. 2025.