

El uso de la evaluación en un contexto de institucionalización débil: una aproximación desde el *Qualitative Comparative Analysis*

Carla Cordoncillo Acosta¹

carla.cordoncillo.acosta@gmail.com

Institut de Govern i Polítiques Públiques (IGOP – UAB)

Resumen

El auge de las políticas basadas en evidencias ha revivido el interés por los usos instrumentales de las evaluaciones. La evidencia disponible sobre el tema se ha generado, en su mayoría, en países con una cultura de la evaluación consolidada, de forma que se dispone de poca información sobre cómo se alcanzan este tipo de usos en contextos donde la evaluación no está completamente institucionalizada. En esta investigación se utiliza un *Qualitative Comparative Analysis* para identificar cuáles son los factores que explican el uso instrumental de las evaluaciones realizadas por el Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua) en Cataluña, un contexto de institucionalización de la evaluación todavía débil. Estos se contrastan con aquellos encontrados por otros investigadores/as en entornos de institucionalización avanzada para identificar diferencias. Tras constatar la importancia explicativa de los tres tipos de factores recogidos en la literatura (metodológicos, institucionales y relativos al contexto político) y la naturaleza coyuntural del uso de la evidencia, los resultados confirman que los factores que influyen en el uso instrumental de las evaluaciones cambian según el grado de institucionalización de la evaluación. En entornos con una cultura de la evaluación menos desarrollada, cobran importancia factores explicativos orientados, justamente, a paliar esa falta de institucionalización.

Palabras clave: usos instrumentales, institucionalización de la evaluación, *Qualitative Comparative Analysis* (QCA), *policymaking*, cultura de la evaluación

[Este documento está en revisión, de forma que queda prohibida su reproducción, circulación, transformación y, en general, cualquier otra forma de explotación, por cualquier procedimiento, de la totalidad o de cualquier parte, sin autorización expresa y por escrito de la autora]

¹ Estudiante de tercer año de doctorado industrial entre el IGOP (UAB) e Ivàlua, donde trabaja como evaluadora de políticas públicas desde hace dos años y medio.

Introducción

El auge del concepto “políticas basadas en evidencias” desde principios de los 2000 ha reforzado el papel de la evaluación como instrumento clave para mejorar las políticas públicas y ha revivido el interés por su utilización. De hecho, según Patton (1997) es justamente ese potencial para mejorar las políticas públicas, y por lo tanto de forma indirecta el bienestar de la sociedad, lo que justifica que se preste tanta atención a cómo se utilizan sus resultados. Toda persona evaluadora espera que sus análisis sirvan para informar la toma de decisiones relativas al diseño y la implementación de políticas públicas (Weiss et al., 2005). Sin embargo, aunque este tipo de uso, denominado instrumental, pueda parecer el más evidente, es solamente uno de los usos posibles de una evaluación, y no siempre el más frecuente.

Sí es el tipo de uso más estudiado. Desde mediados de los años 70s, investigadores e investigadoras se han esforzado en identificar los factores que determinan los usos instrumentales. Han identificado decenas de ellos y los han categorizado en tipologías (Balthasar, 2006; Cousins y Leithwood, 1986; Henry y Mark, 2003; Lester y Wilds, 1990; Pattyn y Bouterse, 2020) que han servido de base para numerosos análisis empíricos. La gran mayoría de dichos análisis, no obstante, fueron realizados en los países llamados “pioneros” o “de la primera ola” en relación con la institucionalización de la evaluación (Estados Unidos, Canadá, Gran Bretaña, Suiza y los Países Bajos). Países que desde hace un par de décadas cuentan con una cultura de la evaluación consolidada (Furubo et al., 2002).

Sin embargo, los factores que explican el desarrollo de una cultura de la evaluación en los países pioneros no son los mismos que la explican en los países que empezaron a desarrollarla décadas después, donde la evaluación llegó casi por obligación, como requisito de los fondos europeos o del Banco Mundial (Viñas, 2009). Según Furubo et al. (2002), el desarrollo de una cultura de evaluación requiere, primero, su institucionalización, y posteriormente, su generalización. Pero a día de hoy, ese primer paso que constituye la institucionalización sigue sin ser una realidad en muchos países, como por ejemplo España, Italia o Irlanda (Jacob et al., 2015).

Por lo tanto, ¿cabe esperar que los factores que influyen en el uso instrumental de las evaluaciones en contextos donde la institucionalización de la evaluación es todavía débil coincidan con los identificados en los países pioneros? Y si no coinciden, ¿qué factores explican el uso instrumental de las evaluaciones en países donde su institucionalización es todavía débil? Este artículo pretende responder a estas dos preguntas a partir de los siguientes pasos: primero, se identifican los factores que, de acuerdo con la evidencia empírica disponible, afectan al uso instrumental de las evaluaciones en entornos de institucionalización avanzada. Segundo, se describen los contextos de institucionalización débil y se selecciona un caso de estudio paradigmático: Ivàlua, en Cataluña. Tercero, se utiliza el *Qualitative Comparative Analysis* (QCA) para identificar las combinaciones de factores que explican el uso instrumental en el caso de estudio, y por último, se contrastan los resultados con los identificados en contextos de institucionalización madura para extraer conclusiones y apuntar futuras líneas de investigación.

Marco analítico: factores que influyen el uso de las evaluaciones en contextos de institucionalización avanzada

Según Patton (1997), lo que justifica el interés por el uso de las evaluaciones es su potencial para mejorar las políticas públicas. Su naturaleza práctica y analítica explica su capacidad para generar conocimiento orientado a fundamentar la toma de decisiones de política pública. Este es el uso que, por lo general, las personas evaluadoras esperan que se dé a sus análisis (Weiss et al., 2005). Y también el uso que está implícito detrás de la concepción de la evaluación como herramienta clave para la promoción de políticas basadas (o informadas) en evidencias.

Sin embargo, no es el único uso posible de las evaluaciones, ni de hecho el más habitual (Pattyn y Bouterse, 2020). La literatura académica sobre el uso del conocimiento en el ámbito de las políticas públicas ha alcanzado cierto consenso alrededor de 4 tipos de usos de las evaluaciones: los usos instrumentales, conceptuales, simbólicos y vinculados al proceso de evaluación. Dicha tipología data de finales de los años 70, la llamada “edad dorada” de la investigación sobre el uso de las evaluaciones (Henry y Mark, 2003), y por lo tanto se atribuye a algunos de los autores que sentaron las bases de la investigación sobre utilización de la evaluación (Alkin y Daillak, 1979; Knorr, 1977; Patton et al., 1977; Weiss, 1979). Posteriormente han surgido propuestas alternativas conceptualmente más complejas (Kirkhart, 2000; Mark y Henry, 2004). Sin embargo, “la mayoría de los estudios recientes trabajan con la tipología convencional de usos de la evaluación, a pesar de reconocer sus debilidades” (Ledermann, 2012, p. 161).

- Los *usos instrumentales* son aquellos en los cuales los resultados y recomendaciones de la evaluación son utilizados para direccionar la toma de decisiones, generalmente en relación al objeto de evaluación (Alkin y Taut, 2002; Weiss et al., 2005). Por lo general, las decisiones refieren al cierre o la continuidad del programa o la política, a su reducción o su expansión. Por lo tanto, cuando se habla de “políticas basadas en evidencias” se está asumiendo un uso instrumental del conocimiento científico.
- Los *usos conceptuales* son aquellos que llevan a un mejor entendimiento del objeto de evaluación (Ledermann, 2012). Aunque no derivan en decisiones o acciones concretas, sí recogen todos aquellos insumos que facilitan una mejor comprensión del programa o la política (Alkin y Taut, 2002).
- Haciendo referencia a Owen (1999), Alkin y Taut (2002) distinguen dos tipos de *usos simbólicos* de una evaluación: cuando esta se utiliza para justificar una decisión tomada previamente (uso legitimador), y cuando se utiliza para mejorar la reputación de quien gestiona la política o el programa o toma las decisiones.
- Patton (1997) propuso añadir a la tipología aquellos *usos derivados del proceso de evaluación* y, aunque hay cierto debate respecto a si son equivalentes a los demás tipos de uso o no (Weiss et al., 2005), la literatura reciente los reconoce como usos y subraya la importancia de tenerlos en cuenta.

Más allá de identificar los diferentes usos posibles de una evaluación, la literatura académica se ha enfocado en identificar los factores determinantes de estos distintos usos y organizarlos en categorías. La primera categorización y seguramente también la más utilizada es la creada por Cousins y Leithwood (1986), quienes distinguen entre: (a)

factores relacionados con el proceso de implementación de la evaluación y (b) factores relacionados con la definición de políticas públicas (*decision or policy setting*). Entre los primeros identifican: la calidad de la evaluación, la credibilidad, la relevancia, la calidad de la comunicación, los resultados y la oportunidad del momento en que sea realiza el análisis. Entre los segundos incluyen: las necesidades de información, las características de la decisión a tomar, el clima político, la información en competencia, las características personales y el compromiso con la evaluación.

Poco tiempo después, Lester y Wilds (1990) proponen una categorización alternativa de lo que denominan “obstáculos a la utilización del conocimiento”. Tal como muestra la Tabla 1, distinguen entre factores técnicos (metodológicos), factores contextuales (políticos) y factores burocráticos (psicológicos). En realidad, las dos primeras categorías son relativamente coincidentes con las de Cousins y Leithwood (1986). Sin embargo, añaden una tercera, que recoge todos los factores individuales (relativos al perfil, el cargo, las capacidades y los intereses personales de los *policymakers*), que ellos incluían en la segunda categoría, así como distintos aspectos vinculados a la cultura organizacional, como el estilo de toma de decisiones (Webber, 1984), la claridad de los objetivos de política pública (Rein y White, 1977) o el nivel de implicación de los tomadores de decisiones en el análisis de las políticas (Lehne y Fisk, 1974), que Cousins y Leithwood (1986) no contemplaron.

Tabla 1: Categorizaciones de factores que influyen en el uso de las evaluaciones

Cousins y Leithwood (1986)	Lester y Wilds (1990)	Balthasar (2006)	Pattyn y Bouterse (2020)
Definición de políticas públicas	Factores contextuales (políticos)	Entorno de la evaluación	Contexto político
			Momento / oportunidad
Implementación de la evaluación	Factores burocráticos (psicológicos)	Contexto institucional	Contacto entre evaluador/a y <i>policy maker</i>
			Características del evaluador/a
	Factores técnicos (metodológicos)	Proceso de evaluación	Características de la administración pública
			Características de la evaluación

Fuente: elaboración propia a partir de las fuentes citadas.

A lo largo de los años 90, otros autores y autoras ampliaron la lista de factores que influyen en el uso de las evaluaciones sin llegar a proponer categorizaciones (Huberman y Gather, 1991; Shulha y Cousins, 1997). Más recientemente tanto Balthasar (2006) como Pattyn y Bouterse (2020) revisaron esa larga lista de factores y propusieron nuevas formas de ordenarlos que, a pesar de utilizar una nueva terminología u optar por categorías más específicas, conceptualmente no difieren mucho de la propuesta de Lester y Wilds (1990).

La investigación sobre los usos de la evaluación, sin embargo, no solamente se centró en la identificación y ordenación de los factores que explican la utilización del conocimiento científico para la toma de decisiones de política pública. La llamada edad dorada también

cubrió un intenso debate de tipo epistemológico sobre la importancia de los factores contextuales a la hora de explicar los usos de las evaluaciones, que “enfrentó” a constructivistas (Guba y Lincoln, 1989; Shulha y Cousins, 1997) con los defensores del carácter político del *policymaking*, quienes defienden que es la actividad política lo que influencia el contexto en el que se producen las evaluaciones, y ese contexto, a su vez, influye en el uso que se acaba realizando de sus resultados y recomendaciones (Greene, 1990; Parkhurst, 2017; Weiss, 1988).

El hecho de que el grueso de la evidencia teórica y empírica sobre los usos de la evaluación esté rigurosamente sistematizada en artículos de referencia que cubren esos 30 años de producción (Cousins y Leithwood, 1986; Johnson et al., 2009; Shulha y Cousins, 1997), es una muestra de la centralidad del tema en el ámbito de las políticas públicas. Esta situación facilita enormemente el análisis de los lugares -y por lo tanto los contextos- desde los cuales se ha producido esta literatura. Así, es posible afirmar que la inmensa mayoría de la evidencia empírica sobre usos de las evaluaciones ha sido producida en Estados Unidos, Canadá, y más recientemente en el norte de Europa (Gran Bretaña, Suiza y Holanda principalmente). Es decir, en países con una larga y consolidada cultura de la evaluación, considerados además “pioneros” en el uso de las evaluaciones (Viñas, 2009).

El concepto de “cultura de la evaluación” se refiere al compromiso con el papel de la evaluación en los procesos de toma de decisiones (Owen y McDonald, 1999). Aunque no son exactamente sinónimos, en la literatura se ha utilizado a menudo para hacer referencia a la institucionalización de la evaluación (Meyer et al., 2020), que constituye el paso previo y necesario para el desarrollo de una cultura de la evaluación. De acuerdo con Meyer et al. (2020) la institucionalización de la evaluación se produce en tres esferas: la política (asegurando el uso de la evaluación como una herramienta de gobernanza generalizada), la social (extendiendo la evaluación a las organizaciones de la sociedad civil y promoviendo que sus resultados informen los debates públicos) y la profesional (vinculando las demandas públicas a una red de profesionales especializados).

Tanto la versión original del *International Atlas of Evaluation* (Furubo et al., 2002) como la versión actualizada (Jacob et al., 2015) muestran la existencia de diferencias profundas en el nivel de institucionalización de la evaluación entre los países de la OCDE. Para explicar esas importantes diferencias, algunos estudios se han centrado en explorar el efecto de la cultura política nacional (Barbier & Hawkins, 2012) o de la cultura organizacional de las administraciones públicas -legislación, antecedentes históricos, conexión con profesiones específicas y evolución de normas y prácticas- (Wilkins, 2012) en la práctica de la evaluación. De acuerdo con Viñas (2009), tres factores explican el desarrollo temprano de la evaluación en los países denominados “de la primera ola”: (1) la predisposición, entendida como normas sociales y actitudes críticas y reflexivas en la administración pública que favorecen la innovación (Williams et al., 2002); (2) la familiaridad del sistema administrativo con el pensamiento científico-social (Derlien, 1990); y (3) “una necesidad satisfecha”, pues a medida que aumenta el gasto en políticas públicas aumenta el interés por monitorizar cómo se gasta el dinero y, por lo tanto, por evaluar y utilizar las evaluaciones (Furubo y Sandahl, 2002). En cambio, en los países que empezaron a evaluar en los 90’ (como España), la evaluación llegó casi por obligación, como requisito de los fondos europeos o del Banco Mundial (Viñas, 2009).

Siendo así, **¿cabe esperar que los factores que influyen el uso instrumental de las evaluaciones en contextos donde la institucionalización de la evaluación es todavía débil coincidan con los identificados en los países pioneros?** Probablemente no, pues los factores que Hornby y Perera (2002) identifican como obstáculos al uso instrumental del conocimiento en los países en vías de desarrollo en su investigación en Sri Lanka - falta de mecanismos de gestión del rendimiento y de monitoreo de la calidad de los servicios, y evaluaciones a demanda-, refieren claramente al nivel de institucionalización de la evaluación. Entonces, **¿qué factores explican el uso instrumental de las evaluaciones en países con una institucionalización de la evaluación todavía en construcción?** Este artículo pretende responder esta pregunta a partir de los siguientes pasos: primero, se han identificado en este apartado los factores que, de acuerdo con la evidencia empírica disponible, afectan el uso instrumental de las evaluaciones en entornos de institucionalización avanzada. Segundo, se describen los contextos de institucionalización débil y se selecciona un caso de estudio paradigmático: Cataluña. Tercero, se utiliza el *Qualitative Comparative Analysis* (QCA) para identificar las combinaciones de factores que explican el uso instrumental en el caso de estudio. Por último, se contrastan los resultados con los identificados en contextos de elevada institucionalización de la evaluación para extraer conclusiones y apuntar futuras líneas de investigación.

La institucionalización de la evaluación en España

En su emblemático *International Atlas of Evaluation*, (Furubo et al., 2002:303) concluían que, “a pesar de empezar a multiplicarse, la evaluación en España se encontraba todavía en una etapa inicial de desarrollo”. Un poco más de 10 años después Jacob et al. (2015) actualizaron el diagnóstico manteniendo los nueve indicadores propuestos por Furubo y sus compañeros. De los 19 países de la OCDE analizados, España se encontraba en 2011 en la posición número 17 en cuanto a institucionalización de la evaluación, solamente seguida por Italia e Irlanda.

Viñas (2009) se basa en el modelo de Furubo y Sandahl (2002) para identificar los factores que explican el desarrollo tardío de la evaluación en España: (1) la escasa inversión en programas sociales en los 70s; (2) una débil tradición de investigación social aplicada; (3) la formación orientada al derecho administrativo de la élite administrativa del país; (4) la falta de incentivos a la evaluación; (5) el sistema parlamentario basado en la disciplina de partidos, que limita su función controladora; y (6) unas estructuras de auditoría pública enfocadas exclusivamente al control del gasto público. Sin embargo, de entre los países analizados por Jacob et al. (2015), también se encontraba entre los 3 que más había mejorado entre 2001 y 2011 (pasando de una puntuación de 5 a una de 11.3).

De acuerdo con Bustelo (2020), en ese período algunos sectores empezaron a realizar una práctica más sistemática de la evaluación, se desarrollaron las primeras formaciones especializadas sobre evaluación y se publicaron las primeras tesis doctorales vinculadas con el tema. La principal apuesta en términos de institucionalización fue la creación de la Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (AEVAL), a nivel nacional, y del Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua), a nivel autonómico, ambas en 2006. Sin embargo, los efectos de la crisis

económica de 2008 frenaron rápidamente el proceso de institucionalización y la década del 2010 se caracterizó por un estancamiento de la práctica de la evaluación y una disminución de su presupuesto, que culminó con la disolución de la AEVAL en 2017.

De acuerdo con algunos autores esta falta de liderazgo estatal de la institucionalización de la evaluación en España habría contribuido a desarrollar una dimensión vertical o territorial del sistema de evaluación español (García Sánchez, 2015). A lo largo de esa misma década distintas comunidades autónomas empezaron a desarrollar mecanismos de institucionalización de la evaluación. Sin embargo, las diferencias entre comunidades autónomas siguen siendo relevantes e Ivàlua (en Cataluña) se mantiene como la única agencia gubernamental de evaluación de tipo generalista, que trabaja desde hace más de una década por la institucionalización de la evaluación, lo cual la ha afianzado como referente de la evaluación, no solamente en la comunidad autónoma, sino también en el resto del Estado (Lázaro, 2015). Todo ello convierte a Cataluña en un interesante caso de estudio, pues dispone de un nivel de institucionalización de la evaluación suficiente como para disponer de información sistematizada, en un contexto todavía de desarrollo de la cultura de la evaluación, de acuerdo con los estándares de la OCDE (Barberà et al., 2020).

Casos de estudio

En este artículo se analizan 20 evaluaciones realizadas por el Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua) entre 2016 y 2020. Se trata del total de evaluaciones entregadas por el Instituto -a las administraciones que las encargaron- en ese período de tiempo, de forma que no se ha realizado ningún tipo de selección. La limitación temporal pretende asegurar que las personas entrevistadas y encuestadas recordaban el proceso de evaluación, así como sus usos posteriores. Adicionalmente, existía la voluntad expresa de analizar un número intermedio de casos para poder aprovechar al máximo el potencial del método (*Qualitative Comparative Analysis*), que funciona especialmente bien con entre 10 y 40 casos (Ragin, 2000; Schneider y Wagemann, 2012). El hecho de que todas las evaluaciones fueran realizadas por la misma entidad (una agencia generalista) facilita mantener constantes condiciones relativas a las características de quien evalúa, la formalidad de su relación con quién encarga la evaluación y estándares de calidad de los informes, lo cual a su vez permite abrir el análisis a distintos ámbitos sectoriales para que sea más generalizable.

La información sobre cada una de las evaluaciones se ha obtenido a partir de tres fuentes. La primera son los informes de resultados de cada una de las evaluaciones, entregados a las administraciones que los encargaron y publicados en la web de Ivàlua. La segunda son entrevistas semi-estructuradas realizadas por el equipo de Ivàlua a las personas que encargaron cada una de las evaluaciones; en su mayoría cargos directivos de distintas administraciones catalanas, aunque también participaron algunas personas técnicas que están a cargo de las políticas en cuestión. Dado que las entrevistas ya estaban hechas en el momento de iniciar esta investigación, se accedió directamente a las transcripciones y grabaciones. La tercera es una encuesta realizada al/la coordinador/a de cada uno de los equipos de evaluación.

Metodología

El *Qualitative Comparative Analysis* (en adelante QCA) es un método de investigación basado en el análisis de casos, entendidos como combinaciones de condiciones. Se atribuye su origen a Charles Ragin, quien en su libro seminal *The Comparative Method* (1987:84) estableció que el QCA "entendía los conjuntos como configuraciones de partes". Es decir, es un método basado en la lógica de conjuntos (*set-theoretic method*) y como tal explora las condiciones específicas en las que se produce un determinado resultado (Mahoney, 2010). Así, a partir de un procedimiento analítico estructurado y un algoritmo (aplicado a partir de un software), el QCA permite comparar casos para identificar condiciones necesarias y suficientes para un determinado resultado (Mello, 2021). Su principal fortaleza es su capacidad de considerar tanto la ecuifinalidad como la causalidad coyuntural (Rihoux y Ragin, 2008). La primera implica que se puede llegar al resultado a partir de diferentes configuraciones (combinaciones de condiciones). La segunda, que esas combinaciones de condiciones pueden ser (conjuntamente) necesarias y/o suficientes para el resultado, aunque esas mismas condiciones no lo sean individualmente. Esta forma de entender la realidad resulta particularmente adecuada para analizar la forma como los diferentes factores identificados influyen el uso instrumental de la evaluación.

Definición de las condiciones

La selección de condiciones se ha realizado estableciendo un diálogo entre la literatura expuesta y los factores explicativos del uso que emergieron durante las entrevistas realizadas. Siguiendo las indicaciones de Pattyn y Bouterse (2020), no se han tenido en cuenta aquellos factores que remiten a aspectos individuales (ya sea de la persona evaluadora, de quien encarga la evaluación o de quien toma las decisiones), pues las decisiones de política pública raramente recaen en una única persona (Weiss, 1998).

Los factores técnicos o metodológicos refieren principalmente a la rigurosidad del análisis y calidad del informe de evaluación (Balthasar, 2006; Johnson et al., 2009; Pattyn y Bouterse, 2020) y a la medida en que este aporta información nueva. Aunque la literatura no vincula directamente esta cuestión con el método de evaluación, los *Randomized Controlled Trials* (RCTs) han sido consideradas desde principios de siglo, el *gold standard* de la evaluación (Deaton y Cartwright, 2017). Así, aunque el término "evidencias" puede referir a muchos tipos de conocimiento, el llamado movimiento de las políticas basadas en evidencias ha considerado la evidencia científica como la más adecuada para informar el diseño de las políticas públicas (Parkhurst, 2017: 16). Con la construcción de "jerarquías de evidencias", ha colocado los métodos experimentales (o en su defecto los cuasiexperimentales) en la cúspide de la jerarquía (Parkhurst, 2017), desprestigiando los demás métodos de evaluación. Por lo tanto, cabría esperar que el método de evaluación influya en su potencial de uso, y por esta razón la primera condición tenida en cuenta es si el caso corresponde a una evaluación de impacto (realizada a partir de métodos experimentales o cuasiexperimentales) o no (IMP).

Balthasar (2006) define los factores burocráticos o relativos al contexto institucional como aquellos que marcan la distancia entre evaluadores/as y evaluados/as. Sin embargo,

ninguno de los 6 factores que propone resulta relevante en este caso por falta de variación, pues casi todas las evaluaciones realizadas en Cataluña en ese periodo fueron demandas de administraciones públicas que no cuentan con una unidad propia de evaluación y por lo tanto contratan a Ivàlua directamente para que realice una evaluación externa. Pattyn y Bouterse (2020), en cambio, definen la interacción entre evaluadores y *policymakers* con base en la implicación de los segundos en el proceso de evaluación: la frecuencia de su contacto con el equipo evaluador y su implicación en la discusión de los resultados. Ellas deciden no considerar estos factores en su análisis porque no encuentran suficiente variación entre casos, pero entre las evaluaciones realizadas por Ivàlua sí es posible identificar diferencias sustanciales. Por esta razón, la frecuencia y la calidad de la relación entre equipo evaluador y *policymaker* (REL) es la segunda condición considerada.

De acuerdo con (Lester y Wilds, 1990: 316), “la naturaleza del problema [también] afecta la utilización, pues algunos problemas son más intratables que otros”. Esta visión encaja con la de Ledermann, (2012), quien señala la conflictividad política alrededor del objeto de evaluación y la presión política por un cambio como factores contextuales (de tipo político) explicativos del uso. Así pues, se ha seleccionado la sensibilidad política del objeto de evaluación (SENS) como tercer factor explicativo.

De esta forma, se propone considerar un factor relativo a cada una de las grandes categorías propuestas tanto por Lester y Wilds (1990) como por Balthasar (2006). No obstante, en buena parte de las entrevistas realizadas, los *policymakers* otorgaban de forma directa o indirecta la mayor parte del peso explicativo del uso instrumental a un factor no identificado directamente en la literatura: la presentación de los resultados de la evaluación (por parte del equipo evaluador) delante del cargo político (POL) encargado de esa unidad. Es por esta razón, que decidió incluirse esta cuarta condición en el análisis.

Así pues, el artículo analiza de qué manera estos cuatro factores explican los usos instrumentales (OUT) de las evaluaciones en un contexto de bajo nivel de institucionalización de la evaluación. De entre todos los tipos de usos mencionados en el marco analítico se eligen los instrumentales porque la mayor parte de la evidencia empírica generada sobre usos se centra en los de tipo instrumental y el objetivo de este artículo es enriquecer esa evidencia acumulada con un análisis generado en un contexto distinto. Adicionalmente, se ha comentado que los usos instrumentales se encuentran detrás de la idea de políticas basadas en evidencias y con la inclusión de la condición de tipo metodológico quiere relacionarse ambos debates. En ese sentido, resulta relevante mencionar que casi la mitad de las evaluaciones analizadas (9 de 20) han tenido un uso instrumental.

Calibrage

El QCA requiere transformar la información empírica disponible sobre cada uno de los casos en niveles pertenencia al conjunto que representa cada una de las condiciones; este proceso se denomina “calibrage” (Schneider y Wagemann, 2012). En este caso, se optó por un *fuzzy-set* QCA -es decir, un tipo de QCA que permite asignar a cada caso membresías con valor de 0 a 1- para poder captar mejor las diferencias entre casos.

Para transformar la información cualitativa relativa a los usos instrumentales (OUT) en valores *fuzzy-set* se siguió el procedimiento definido por Basurto y Speer (2012). Así, el umbral de 0.5 separa aquellas evaluaciones que tuvieron usos instrumentales de las que no los obtuvieron. Las membresías por encima del umbral se otorgaron en función del número de usos instrumentales y de sus implicaciones presupuestarias. Los valores por debajo se definieron en función del número de usos conceptuales. Las condiciones relativas a la relación entre equipo evaluador y *policymaker* (REL) y a la sensibilidad política del objeto de evaluación (POL) se habían recogido a partir de escalas *Likert* de 4 puntos que se transformaron directamente en membresías de 0, 0.3, 0.7 i 1. Las condiciones de evaluación de impacto (IMP) y haber presentado los resultados ante un cargo político (POL) se abordaron como *crisp* por su naturaleza dicotómica. En los materiales adicionales se pueden consultar los detalles sobre el proceso de calibrage, así como la tabla con las cuatro condiciones y el resultado calibrados para las 20 evaluaciones (*raw data*).

Resultados

El primer paso en el análisis de un QCA es el estudio de las relaciones de necesidad. Las condiciones *necesarias* son aquellas que siempre están presentes cuando el resultado de interés (en este caso el uso instrumental) se produce. Es decir, son pre-requisitos para la obtención del resultado. La Tabla 2 muestra los principales indicadores a partir de los cuales identificar relaciones de necesidad para cada una de las condiciones analizadas.

Tabla 2. Análisis de condiciones necesarias para el uso instrumental

Condición	<i>Consistency</i>	<i>Coverage</i>	RoN	Condición	<i>Consistency</i>	<i>Coverage</i>	RoN
SENS	0.691	0.691	0.774	~SENS	0.557	0.254	0.664
REL	0.814	0.675	0.686	~REL	0.474	0.554	0.760
POL	0.876	0.531	0.348	~POL	0.124	0.300	0.851
IMP	0.629	0.555	0.647	~IMP	0.371	0.400	0.671

Nota: El símbolo "~" indica la negación de una condición. SENS = objeto de evaluación políticamente sensible; REL = relación buena y frecuente entre equipo de evaluación y *policymaker*; POL = resultados de la evaluación presentados ante cargo política; IMP = evaluación de impacto. RoN = *Relevance of necessity*.

Ninguna de las condiciones supera el umbral de *consistency* de 0.9, necesario para ser considerada una condición necesaria (Schneider y Wagemann, 2012). E igual que en el análisis que realizan Pattyn y Bouterse (2020), tampoco se identifican condiciones necesarias para el no-uso o la ausencia de uso instrumental. Sin embargo, expertos y expertas en QCA reconocen que cuando se analizan fenómenos sociales es muy difícil encontrar condiciones necesarias; por lo que este umbral debe definirse en cada caso teniendo en cuenta el conocimiento sobre el tema, los casos y la interpretación de los datos empíricos (Mello, 2021). En este caso, la condición POL (presentación de los resultados de la evaluación frente al nivel político) tiene una *consistency* de 0.876, muy

cercana al umbral, porque hay un único caso que no presenta la condición y sí presenta el resultado. Se trata de un piloto sobre un aspecto muy concreto de la implementación de una política educativa cuyo impacto se había evaluado y presentado ante cargos políticos unos meses antes. Lo cual apuntaría que POL podría ser considerada una condición casi-necesaria para el uso instrumental de la evaluación. Sin embargo, es una condición trivial, pues tanto su *coverage* como su RoN son muy bajos, de forma que por sí sola es una condición poco explicativa del resultado.

El hecho de que abordar el impacto de la política no sea una condición necesaria para el uso instrumental de la evaluación demuestra que, a pesar de lo que defienden las definiciones más restrictivas de políticas basadas en evidencias (Baron, 2018), otros tipos de evaluación (de necesidades, diseño e implementación), no basados en métodos experimentales, también pueden derivar en decisiones de política pública. Aún y así, de los 9 casos que derivaron en un uso instrumental, 6 son evaluaciones de impacto (o económicas), lo cual parece corroborar que la legitimidad del método de evaluación (o tal como lo planteó Bozeman (1986), su credibilidad) es un factor que influye en su uso.

Análisis de la truth table

La *truth table* es el elemento analítico central del QCA y sirve para identificar condiciones y/o combinaciones de condiciones *suficientes* para el resultado. La Tabla 3 muestra la *truth table* para el resultado uso instrumental (OUT) y las condiciones explicativas (SENS, REL, POL, IMP)². Para cada fila la *truth table* proporciona el número de casos que presentan esa configuración (n), la medida en qué es suficiente para la obtención del resultado (Incl) y una medida adicional de consistencia (PRI = *Proportional Reduction in Inconsistency*) que permite identificar contradicciones lógicas (Mello, 2021). Para su minimización se ha fijado un nivel de consistencia de 0.8 (Mello, 2021), de forma que solo se han incluido las tres primeras filas (aquellas que sobrepasan el umbral. La Tabla 4 muestra la solución conservadora de acuerdo con la notación propuesta por Ragin y Fiss (2008), es decir sin tener en cuenta los *logical reminders* en el análisis.

² Dado que el modelo incluye 4 condiciones la *truth table* tiene $2^4 = 16$ filas de posibles combinaciones. Las filas para las cuales no se dispone de casos empíricos se denominan *logical reminders*.

Tabla 3. *Truth table* para el resultado uso instrumental

SENS	REL	POL	IMP	OUT	n	Incl	PRI	Casos
1	0	1	1	1	1	1.000	1.000	5
1	1	1	1	1	3	1.000	1.000	3, 4, 6
1	1	1	0	1	2	0.862	0.750	1, 10
0	1	1	0	0	1	0.773	0.286	8
0	1	1	1	0	4	0.676	0.455	11, 16, 17, 19
0	0	1	0	0	2	0.654	0.182	2, 9
1	0	1	0	0	3	0.636	0.400	12, 13, 20
0	1	0	1	0	2	0.588	0.462	7, 18
0	0	0	1	0	1	0.462	0.222	14
1	0	0	0	0	1	0.000	0.000	15
0	0	0	0	?	0	-	-	-
0	0	1	1	?	0	-	-	-
0	1	0	0	?	0	-	-	-
1	0	0	1	?	0	-	-	-
1	1	0	0	?	0	-	-	-
1	1	0	1	?	0	-	-	-

Nota: SENS = objeto de evaluación políticamente sensible; REL = relación buena y frecuente entre equipo de evaluación y *policymaker*; POL = resultados de la evaluación presentados ante cargo política; IMP = evaluación de impacto.

La primera configuración de la solución combina un objeto de evaluación políticamente sensible, analizado de forma colaborativa entre equipo evaluador y *policymaker* (relación buena y frecuente) y cuyos resultados se han presentado ante un cargo político (SENS*REL*POL). La segunda configuración de la solución combina un objeto de evaluación políticamente sensible, analizado a partir de una metodología de evaluación experimental o cuasiexperimental y presentado ante el nivel político (SENS*POL*IMP). De las 9 evaluaciones con un uso instrumental 6 quedan cubiertas por esta solución. Por esta razón, aunque la *consistency* del modelo es muy alta, su cobertura es de 0.644. De hecho, 3 de los casos explicados por el modelo quedan cubiertos por ambas configuraciones, pues presentan las 4 condiciones, lo cual evidencia, en primer lugar, que todos los tipos de factores que pueden influenciar el uso son de alguna forma relevantes, y, en segundo lugar, que es su combinación lo que conduce al uso instrumental.

Tabla 4. Configuraciones para un uso instrumental

	Conf. (path) 1	Conf. (path) 2
Sensibilidad política del objeto de evaluación	●	●
Relación buena y frecuente entre equipo evaluador y <i>policymaker</i>	●	
Presentación de los resultados al nivel político	●	●
Evaluación de impacto		●
<i>Consistency</i>	0.938	1.000
PRI	0.900	1.000
<i>Raw coverage</i>	0.604	0.396
<i>Unique coverage</i>	0.248	0.040
Casos cubiertos	1,10; 3, 4, 6	5; 3; 4; 6
<i>Solution consistency</i>	0.942	
<i>Solution PRI</i>	0.909	
<i>Solution coverage</i>	0.644	
Modelo	M1 (2)	

Nota: los puntos negros indican presencia de la condición y los puntos cruzados ausencia.

La Tabla 5 muestra la solución conservadora para la ausencia de uso instrumental. El modelo incluye tres configuraciones que, conjuntamente, cubren 5 casos. La primera combina un objeto de evaluación poco sensible políticamente, abordado a partir de una metodología no experimental y presentado ante el nivel político (\sim SENS*POL* \sim IMP). La segunda, las evaluaciones de impacto de políticas que no son sensibles, cuyo proceso de realización no ha pasado por una buena y fluida relación entre evaluadores y evaluados y que no se han llegado a presentar ante ningún cargo político (\sim SENS* \sim REL* \sim POL*IMP). La tercera, las evaluaciones que no son de impacto sobre temas políticamente sensibles en las que la relación no ha sido buena entre evaluadores y contraparte y tampoco se han presentado ante el nivel político (SENS* \sim REL* \sim POL* \sim IMP).

Tabla 5: Configuraciones para la ausencia de uso instrumental

	Conf. (path) 3	Conf. (path) 4	Conf. (path) 5
Sensibilidad política del objeto de evaluación	⊗	⊗	●
Relación buena y frecuente entre equipo evaluador y <i>policymaker</i>		⊗	⊗
Presentación de los resultados al nivel político	●	⊗	⊗
Evaluación de impacto	⊗	●	⊗
<i>Consistency</i>	0.848	0.846	1.000
PRI	0.722	0.778	1.000
<i>Raw coverage</i>	0.272	0.107	0.068
<i>Unique coverage</i>	0.146	0.107	0.068
Casos cubiertos	2, 9; 8	14	15
<i>Solution consistency</i>		0.868	
<i>Solution PRI</i>		0.794	
<i>Solution coverage</i>		0.447	
Modelo		M1 (3)	

Nota: los puntos negros indican presencia de la condición y los puntos cruzados ausencia.

Discusión

Para dar respuesta a las preguntas de investigación estos resultados deben entenderse a la luz de la literatura presentada. Lo primero que llama la atención es que casi la mitad de las evaluaciones analizadas (9 de 20) han tenido un uso instrumental, una cifra elevada si se tiene en cuenta que, en un contexto de cultura de la evaluación consolidada, Pattyn y Bouterse (2020) identificaron 5 evaluaciones con un uso instrumental sobre un total de 18. Parecería, por lo tanto, que la dificultad de alcanzar usos instrumentales no se agrava en contextos donde la evaluación está menos institucionalizada. Adicionalmente, de acuerdo con ambas soluciones, los usos instrumentales se confirman como un fenómeno multicausal: los tres tipos de factores (metodológicos, institucionales y relativos al contexto político) son relevantes, pero es su combinación (su interacción) lo que conduce al uso instrumental, también en contextos de baja institucionalización de la evaluación.

En esa combinación, emerge una condición casi-necesaria no identificada en la literatura: la presentación de los resultados de la evaluación ante cargos políticos. La existencia de mecanismos institucionales que garantizan la diseminación y el uso de las evaluaciones por parte del gobierno y del Parlamento es un elemento de institucionalización (Jacob et al., 2015) avanzada, que raramente se da en contextos como Cataluña donde todavía se están desarrollando los mecanismos de institucionalización dentro de la administración pública. La presencia de la condición POL en todas las evaluaciones que derivaron en un uso instrumental excepto una (por motivos previamente expuestos) parece apuntar que,

en ausencia de estructuras formales de institucionalización orientadas a acercar los resultados de las evaluaciones a los decisores de política pública, se crean estructuras informales con una función similar.

Pero ¿por qué es tan importante ese espacio de disseminación, formal o informal? Una posible explicación la podemos encontrar en la idea de las “dos comunidades” de Caplan (1979), según la cual académicos/as -o en este caso evaluadores/as- y *policymakers* tienen formas de actuar y pensar distintas. Así, la presentación y discusión de los resultados de la evaluación constituye un espacio de diálogo que permite acercar ambas comunidades y superar un obstáculo importante para el uso instrumental: las barreras de lenguaje (Mead, 2015). En ese sentido, Jacob (2006) considera la hibridación de las culturas y prácticas administrativas y académicas una condición fundamental para el desarrollo de la evaluación en las administraciones públicas.

En su análisis sobre usos instrumentales en los Países Bajos, un país con una cultura de la evaluación madura, Pattyn y Bouterse (2020) identifican dos condiciones necesarias, aunque triviales: el interés de los *policymakers* en la evaluación y sus resultados y el hecho de que la evaluación se llevara a cabo de forma paralela al diseño de una nueva política. Sin embargo, para el caso de Cataluña ambas presentaban muy poca variabilidad, razón por la cual no se incluyeron en el análisis. Vale la pena mencionar, no obstante, que mientras en los Países Bajos la totalidad de evaluaciones que resultaron en usos instrumentales se habían ejecutado en paralelo al proceso de diseño de una nueva política pública (o rediseño de la política evaluada), de entre las realizadas por Ivàlua solamente dos cumplen esta condición. Esta diferencia podría deberse a una menor planificación de las evaluaciones en contextos de institucionalización débil.

Por otro lado, la relación entre equipo evaluador y *policymakers* (que ellas identifican como constante en todos sus casos) se muestra una condición con amplia variabilidad en Cataluña. El grado de implicación de los *policymakers* en los procesos de evaluación varía según distintos factores (su interés, sus conocimientos técnicos, la complejidad de la evaluación), lo cual se traduce en variaciones importantes tanto en la frecuencia del contacto que mantienen con el equipo evaluador, como en la calidad de su relación. Así, esta condición repetidamente apuntada por la literatura (Marra, 2004; Marsh y Glassick, 1988; Preskill et al., 2003) se confirma como relevante para explicar los usos instrumentales de la evaluación en contextos de institucionalización débil. Uno de los argumentos detrás de la relevancia de esta condición es que, dada la imposibilidad de los *policymakers* de considerar toda la evidencia relevante sobre cada problema de política pública (Cairney & Oliver, 2017), una relación cercana y favorable con el equipo evaluador facilita la priorización de la evidencia generada en el marco de la evaluación por encima de hábitos o intuiciones contrarias.

En ese sentido, resulta interesante que, en la solución presentada en la Tabla 4, la relación cercana entre equipo evaluador y *policymakers* y el uso de métodos experimentales o cuasiexperimentales parecen ser intercambiables. Mientras que el objeto de evaluación políticamente sensible y la presentación de los resultados de la evaluación ante un cargo político están presentes en ambas configuraciones de la solución, estas dos condiciones se alternan (aunque 3 de los casos encajan en ambas configuraciones porque cuentan con

las 4 condiciones). Una posible explicación es que ambas condiciones refuerzan la confianza de los *policymakers* en los resultados de la evaluación.

De acuerdo con Parkhurst (2017), el movimiento de políticas basadas en evidencias y su defensa de que la evidencia nos dice “qué funciona” para alcanzar determinados resultados de política pública y así elegir las soluciones más efectivas para resolver problemas sociales y ahorrar recursos públicos, ha llevado a confundir el rigor de la evidencia con la importancia de la política. Esta confusión puede conllevar asimismo el riesgo de priorizar aquellas políticas que pueden ser evaluadas mediante métodos experimentales o cuasiexperimentales (lo que se ha denominado *issue bias*). Aunque no disponemos de la información necesaria para valorar si la administración pública catalana ha incurrido o no en este tipo de sesgo a la hora de priorizar las políticas a impulsar y por lo tanto también a evaluar, resulta relevante que 11 de las 20 evaluaciones realizadas por Ivàlua en el período 2016-2020 sean evaluaciones de impacto (o económicas, que requieren haber realizado previamente una evaluación de impacto). De forma similar, de las 33 evaluaciones que realizó la institución entre 2008 y 2016, 15 fueron evaluaciones de impacto (Comas & y Vílchez, 2018). Esta proporción constante alrededor del 50% contrasta con el hecho que, en un contexto de cultura de la evaluación madura, de las 18 evaluaciones analizadas por Pattyn y Bouterse (2020) solamente 3 utilizaron métodos experimentales o cuasi-experimentales³.

Muchos factores pueden explicar esta diferencia. De hecho, numerosos autores y autoras han expresado su preocupación alrededor de la idea de *gold standard*, justamente porque los requisitos metodológicos de los RCTs hacen muy difícil su aplicación en determinados ámbitos sectoriales (Cameron et al., 2015; Deaton y Cartwright, 2017; Ebrahim y Rangan, 2010), como podría ser el de los asuntos exteriores, al que pertenecen todos los casos analizados por Pattyn y Bouterse (2020). Sin embargo, resulta interesante constatar que una de las pocas dimensiones en las que España destaca en la actualización de 2015 del Atlas de Evaluación, es justamente la relativa a la proporción de evaluaciones de impacto (en la cual obtiene un 1,5 sobre 2) (Jacob et al., 2015). Esta situación obliga a interrogarse sobre si en contextos con menor cultura de la evaluación se han tendido a priorizar evaluaciones experimentales y por qué. Una posible explicación es que, a falta de presupuesto para evaluaciones, se hayan priorizado aquellas que tienen financiación externa. Otra es que, siendo la práctica de la evaluación escasa, se hayan priorizado las evaluaciones de impacto para mostrar un compromiso -más simbólico que estructural- con la evaluación rigurosa y las políticas basadas en evidencias.

Se ha comentado que las dos configuraciones recogidas en la solución de la Tabla 4 incluyen la condición SENS. Este resultado confirma que, tal como apuntaban Lester y Wilds (1990), el objeto de evaluación es un determinante del uso instrumental. Más concretamente, los autores se fijaban en dos características del problema: su complejidad -proponiendo que contra más difícil de conceptualizar y tratar el problema menos

³ La mayoría de los casos analizados por Pattyn y Bouterse (2020) refieren a evaluaciones de estrategias nacionales, planes de 4 o 5 años de duración o programas que recogen distintos proyectos. Es decir, políticas públicas de gran abasto, lo cual dificulta en buena medida que puedan ser abordadas mediante métodos experimentales o cuasiexperimentales. De todas formas, una de estas evaluaciones utilizó este tipo de métodos para una parte del análisis y otras dos utilizaron las evaluaciones de impacto de algunos de los proyectos incluidos en el programa o la estrategia evaluada como fuentes de información.

probable sería el uso instrumental- y la urgencia de una solución -a mayor urgencia mayor probabilidad de uso-. En este caso se ha analizado una característica que hibrida estas dos con el determinante contextual relativo al nivel de conflictividad alrededor del objeto de evaluación, propuesto por Ledermann (2012). Los resultados apuntan que, en un contexto de institucionalización débil, aquellas evaluaciones que tienen por objeto de análisis políticas públicas que se encuentran en el centro de la agenda política y que por lo tanto reciben mucha más atención mediática y política, tienen más posibilidades de ser utilizadas de forma instrumental. De entre las 9 evaluaciones con usos instrumentales 7 son políticamente sensibles y la falta de esta característica es parte de dos de las configuraciones suficientes para la ausencia de uso instrumental. Este resultado contrasta con el de Pattyn y Bouterse (2020:7), quienes encontraron que -en un entorno con una cultura de la evaluación madura- “el uso de las evaluaciones no se ve perjudicado ni tampoco promovido por la sensibilidad política del objeto de evaluación”.

Aún y así, la solución presentada en la Tabla 5 para la ausencia de usos instrumentales confirma que ni la presentación de resultados (POL) ni el objeto de evaluación políticamente sensible (SENS) son por sí mismas condiciones suficientes para el uso instrumental. De hecho, en ausencia de las demás condiciones, son suficientes para la ausencia de uso. Es decir, solamente en combinación con los demás tipos de factores explicativos, tienen la capacidad de influir positivamente en la utilización del conocimiento. Y este es un resultado muy interesante porque no solo ejemplifica la asimetría causal que permite captar el QCA, sino que además muestra que determinadas condiciones pueden afectar positiva o negativamente el *outcome* en función de cómo interactúan con las demás condiciones (una situación que los métodos estadísticos no permiten captar). Todo ello confirma, como también apuntaban Pattyn y Bouterse (2020b), la naturaleza coyuntural⁴ del uso de la evidencia ya apuntada por Ledermann (2012).

Conclusión

La literatura sobre los usos del conocimiento se ha esforzado en identificar y categorizar los factores que influyen en el uso instrumental de las evaluaciones. La mayoría de la evidencia empírica, no obstante, proviene de los denominados países de la primera ola: países donde la evaluación presenta un alto grado de institucionalización. Así pues, para llenar ese vacío, este artículo se proponía responder a dos preguntas: cuáles son los factores que explican el uso instrumental de las evaluaciones en contextos donde su institucionalización es débil, y si éstos coinciden con los identificados por la literatura en entornos con una cultura de la evaluación madura.

De acuerdo con lo esperado, los resultados confirman dos aspectos. Primero, que los tres tipos de factores identificados por la literatura (metodológicos, institucionales y relativos al contexto político) son relevantes para explicar el uso instrumental de las evaluaciones en contextos donde esta no está completamente institucionalizada. Segundo, que por sí solos ninguno de ellos es suficiente para explicar el uso instrumental de las evaluaciones. Asimismo, este resultado reafirma la naturaleza coyuntural del uso de la evidencia y por

⁴ Relativa a causalidad coyuntural (*conjunctural causation*).

lo tanto la pertinencia de estudiarla a partir de métodos capaces de captar este tipo de relaciones, como es el QCA.

El estudio también ha permitido identificar una serie de diferencias en comparación con aquellos contextos donde la cultura de la evaluación está consolidada. En su artículo sobre el uso y el no-uso de la evaluación en contextos donde está plenamente institucionalizada, Pattyn y Bouterse (2020:2) apuntaban que “las organizaciones con una cultura de la evaluación madura, por definición ya cumplen con muchos de los factores que generalmente se dice promueven el uso de la evaluación”. Efectivamente, ello explica que una primera diferencia relevante sea que las condiciones que muestran una variación suficiente como para ser incluidas en el análisis en ambos contextos sean bien diferentes.

En segundo lugar, mientras Pattyn y Bouterse (2020) concluyeron que la sensibilidad política del objeto de evaluación era una condición que no influía -ni positiva ni negativamente- el uso instrumental de las evaluaciones, en este análisis aparece en las dos configuraciones de la solución. Así, a diferencia de lo que sucede en los contextos donde la evaluación está institucionalizada, en el caso de Ivàlua, en Cataluña, son las evaluaciones políticamente sensibles las que derivan en usos instrumentales. Sería interesante profundizar en el porqué de esta diferencia en futuras investigaciones. La explicación puede estar relacionada con la interacción de la evidencia con los demás factores que influyen en los procesos de toma de decisiones en uno y otro tipo de contextos. Y es que la evidencia es solamente una de las fuentes que informan el proceso político-administrativo (Albæk, 1995), de manera que su influencia nunca se da de forma directa, sino que se hibrida con otros factores como son los intereses, alianzas, redes, creencias, entre otros (Parkhurst, 2017; C. Weiss, 1979). En este sentido, Parkhurst (2017) afirma que la investigación sobre los usos de la evidencia se ha basado mayoritariamente en propuestas normativas sobre cómo y con qué finalidad debería utilizarse el conocimiento en los procesos de diseño y gestión de políticas públicas, y en consecuencia le ha faltado reconocer de forma explícita la naturaleza de la política. El hecho de que en contextos donde la evaluación está menos institucionalizada sean las evaluaciones sobre temas políticamente sensibles las que terminan utilizándose, parece evidenciar esta naturaleza política y contradecir, como sugiere Daviter (2015), la idea implícita en buena parte de la literatura sobre uso del conocimiento de que “la política” restringe el uso y el efecto del conocimiento científico en el *policymaking*.

En tercer lugar, llama la atención que los métodos de evaluación utilizados en ambos contextos son bien distintos. La preeminencia de los métodos experimentales y cuasiexperimentales en Cataluña podría indicar que en contextos donde la evaluación está menos institucionalizada se tienden a priorizar los métodos con mayor credibilidad (porque encabezan las jerarquías de evidencia), como forma de mostrar el compromiso de las administraciones públicas con la evaluación rigurosa. Dada la falta de estructuras para garantizar que las evidencias informen de forma orgánica y sistemática el diseño y la implementación de las políticas públicas, las administraciones podrían estar optando por los métodos con mayor credibilidad a modo de compensación. Sin embargo, esta es una hipótesis que habría que explorar en mayor profundidad en futuras investigaciones, así como el posible *issue bias* derivado de esta situación.

Por último, los resultados apuntan que en ausencia de estructuras formales de institucionalización de la evaluación en el sistema parlamentario y de toma de decisiones, cobran importancia estructuras informales que buscan cumplir con esa misma función. Así lo evidencia la emergencia de la presentación de los resultados de la evaluación ante cargos políticos, un factor no identificado explícitamente en la literatura, como una condición cuasi-necesaria para el uso instrumental de las evaluaciones realizadas por Ivàlua en un contexto de institucionalización débil. Así, sería interesante poder replicar el análisis en otros contextos similares para ver si los resultados coinciden y poder profundizar sobre su generalización.

En todo caso, este estudio muestra que los factores que influyen en el uso instrumental de las evaluaciones cambian según el contexto. En entornos donde la evaluación está menos institucionalizada parecen cobrar importancia condiciones orientadas justamente a paliar esa falta de estructuras y se hace más evidente la naturaleza política de las políticas públicas.

PREPRINT

Bibliografía

Albæk, Erik (1995). Between knowledge and power: Utilization of social science in public policy making. *Policy Sciences*, 28(1), 79-100. <https://doi.org/10.1007/BF01000821>

Alkin, Marvin C. y Daillak, Richard H. (1979). A Study of Evaluation Utilization. *Educational Evaluation and Policy Analysis*, 1(4), 41-49. <https://doi.org/10.3102/01623737001004041>

Alkin, Marvin C. y Taut, Sandy M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1-12. [https://doi.org/10.1016/S0191-491X\(03\)90001-0](https://doi.org/10.1016/S0191-491X(03)90001-0)

Balthasar, Andreas (2006). The Effects of Institutional Design on the Utilization of Evaluation: Evidenced Using Qualitative Comparative Analysis (QCA). *Evaluation*, 12(3), 353-371.

Barberà, Oscar, Doria, Eemilio, Ntutumu-Sanchís, Fernando y Sanchis, Pau (2020). *La institucionalización de la evaluación de políticas públicas. La Comunitat Valenciana en perspectiva comparada*. Publicacions de la Universitat de València. https://puv.uv.es/la-institucionalizacion-de-la-evaluacion-de-politicas-publicas.html?__store=espanyol&__from_store=valencia

Barbier, Jean-Claude y Hawkins, Penny (2012). *Evaluation Cultures: Sense-making in Complex Times* (N.º 2). Transaction Publishers. <https://journalhosting.ucalgary.ca/index.php/cjpe/article/view/30933>

Baron, Jon (2018). A Brief History of Evidence-Based Policy. *The ANNALS of the American Academy of Political and Social Science*, 678(1), 40-50. <https://doi.org/10.1177/0002716218763128>

Basurto, Xavier y Speer, Johanna (2012). Structuring the Calibration of Qualitative Data as Sets for Qualitative Comparative Analysis (QCA). *Field Methods*, 24(2), 155-174. <https://doi.org/10.1177/1525822X11433998>

Bozeman, Barry (1986). The credibility of policy analysis: Between method and use. *Policy Studies Journal*, 14(4), 519-539. <https://doi.org/10.1111/j.1541-0072.1986.tb00358.x>

Bustelo, María (2020). Spain. En W. Meyer, R. Stockmann, & L. Taube, *The Institutionalisation of Evaluation in Europe* (Springer International Publishing). <https://www.springerprofessional.de/en/spain/17885778>

Cairney, Paul y Oliver, Kathryn (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems*, 15(1), 35. <https://doi.org/10.1186/s12961-017-0192-x>

Cameron, Drew, Mishra, Anjini y Brown, Annette (2015). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness*, 8, 1-21. <https://doi.org/10.1080/19439342.2015.1034156>

Caplan, Nathan (1979). The Two-Communities Theory and Knowledge Utilization. *American Behavioral Scientist*, 22(3), 459-470. <https://doi.org/10.1177/000276427902200308>

Comas, Núria y Vílchez, David (2018). *L'ús de les avaluacions d'Ivàlua (2008-2016). Una anàlisi de com les avaluacions enforteixen les polítiques públiques*. Institut Català d'Avaluació de Polítiques Públiques (Ivàlua). https://ivalua.cat/sites/default/files/inline-files/13_12_2018_16_10_18_Informe_Us_avaluacions_1.pdf

Cousins, J. Bradley y Leithwood, Kenneth A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331-364. <https://doi.org/10.2307/1170319>

Daviter, Falk (2015). The political use of knowledge in the policy process. *Policy Sciences*, 48, 491-505. <https://doi.org/10.1007/s11077-015-9232-y>

Deaton, Angus y Cartwright, Nancy (2017). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210. <https://doi.org/10.1016/j.socscimed.2017.12.005>

Derlien, Hans-Ulrich (1990). Genesis and Structure of Evaluation Efforts in Comparative Perspective. En *Program evaluation and the management of government* (Rist, R.C., pp. 147-175). Transaction Publishers. <https://doi.org/10.4324/9780429338380-14>

Ebrahim, Alnoor S. y Rangan, V. Kasturi (2010). The Limits of Nonprofit Impact: A Contingency Framework for Measuring Social Performance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1611810>

Furubo, J. y Sandahl, R. (2002). Introduction. A diffusion perspective on Global Developments in Evaluation. En J. Furubo, R. Rist, & R. Sandahl (Eds.), *International Atlas of Evaluation* (pp. 1-23). Transaction Publishers.

Furubo, Jan-Eric, Rist, Ray y Sandahl, Rolf (2002). *International Atlas of Evaluation*. Transaction Publishers.

García Sánchez, Ester (2015). Sistemas de evaluación en España: Análisis y propuestas. *Studia Politicae*, 34, 117-134.

Greene, Jennifer C. (1990). Technical quality versus user responsiveness in evaluation practice. *Evaluation and Program Planning*, 13(3), 267-274. [https://doi.org/10.1016/0149-7189\(90\)90057-4](https://doi.org/10.1016/0149-7189(90)90057-4)

Guba, Egon G. y Lincoln, Yvonna S. (1989). *Fourth generation evaluation*. Sage Publications.

Henry, Gary y Mark, Melvin (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *The American Journal of Evaluation*, 24, 293-314. [https://doi.org/10.1016/S1098-2140\(03\)00056-0](https://doi.org/10.1016/S1098-2140(03)00056-0)

Hornby, P. y Perera, H. S. R. (2002). A development framework for promoting evidence-based policy action: Drawing on experiences in Sri Lanka. *The International Journal of Health Planning and Management*, 17(2), 165-183. <https://doi.org/10.1002/hpm.660>

- Huberman, Michael y Gather, Monica (1991). De la recherche à la pratique: Éléments de base. *Recherche & formation*, 12(1), 174-176.
- Jacob, Steve (2006). *Institutionnaliser l'évaluation des politiques publiques: Études comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*.
- Jacob, Steve, Speer, Sandra y Furubo, Jan-Eric (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21, 6-31. <https://doi.org/10.1177/1356389014564248>
- Johnson, Kelli, Greenesid, Lija O., Toal, Stacie A., King, Jean A., Lawrenz, Frances y Volkov, Boris (2009). Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377-410. <https://doi.org/10.1177/1098214009341660>
- Kirkhart, Karen E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5-23. <https://doi.org/10.1002/ev.1188>
- Knorr, Karin D. (1977). Policymakers' Use of Social Science Knowledge: Symbolic or Instrumental? En *Using Social Research in Public Policy Making* (Weiss, C.H., pp. 165-182). Lexington Books.
- Lázaro, Blanca (2015). *Estudio comparado sobre institucionalización de la evaluación en Europa y en América Latina* (Estudios N.º 15). Programa EUROsociAL.
- Ledermann, Simone (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159-178. <https://doi.org/10.1177/1098214011411573>
- Lehne, Richard y Fisk, Donald M. (1974). The Impact of Urban Policy Analysis. *Urban Affairs Quarterly*, 10(2), 115-138. <https://doi.org/10.1177/107808747401000201>
- Lester, James. P. y Wilds, Leah J. (1990). The utilization of public policy analysis: A conceptual framework. *Evaluation and Program Planning*, 13(3), 313-319.
- Mark, Melvin y Henry, Gary (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10, 35-57. <https://doi.org/10.1177/1356389004042326>
- Marra, Mita (2004). *The Contribution of Evaluation to Socialization and Externalization of Tacit Knowledge*. <https://doi.org/10.1177/1356389004048278>
- Marsh, David D. y Glassick, Judith M. (1988). Knowledge Utilization in Evaluation Efforts: The Role of Recommendations. *Knowledge*, 9(3), 323-341. <https://doi.org/10.1177/107554708800900301>
- Mead, Lawrence M. (2015). Only connect: Why government often ignores research. *Policy Sciences*, 48(2), 257-272.
- Meyer, Wolfgang, Stockmann, Reinhard y Taube, Lena (2020). *The Institutionalisation of Evaluation Theoretical Background, Analytical Concept and Methods* (pp. 3-34). https://doi.org/10.1007/978-3-030-32284-7_1

Owen, John M. y McDonald, Diane E. (1999). Creating an Evaluation Culture in International Development Cooperation Agencies. *Journal of International Cooperation in Education*, 2(2), 41-53.

Parkhurst, Justin O. (2017). *The politics of evidence: From evidence-based policy to the good governance of evidence*. Routledge. <https://doi.org/10.4324/9781315675008>

Patton, Michael Q. (1997). *Utilization-focused evaluation: The new century text, 3rd ed* (pp. xvi, 431). Sage Publications, Inc.

Patton, Michael Q., Grimes, Patricia S., Guthrie, Kathryn M., Brennan, Nancy J., French, Barbara D. y Blyth, Dale A. (1977). In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research. En *Using social research in public policy making* (Weiss, C.H., pp. 141-164). Lexington Books. <https://eric.ed.gov/?id=ED135938>

Pattyn, Valérie y Bouterse, Marjolein (2020). Explaining use and non-use of policy evaluations in a mature evaluation setting. *Humanities and Social Sciences Communications*, 7(1), 1-9. <https://doi.org/10.1057/s41599-020-00575-y>

Preskill, Hallie, Zuckerman, Barbara y Matthews, Bonya (2003). An Exploratory Study of Process Use: Findings and Implications for Future Research. *The American Journal of Evaluation*, 24(4), 423-442. <https://doi.org/10.1016/j.ameval.2003.09.001>

Ragin, Charles C. (2000). *Fuzzy-Set Social Science*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/F/bo3635786.html>

Ragin, Charles C. y Fiss, Peer C. (2008). Net Effects Analysis Versus Configurational Analysis: An Empirical Demonstration. En *Redesigning Social Inquiry: Set Relations in Social Research* (Ragin). University of Chicago Press.

Rein, Martin y White, Sheldon H. (1977). Policy Research: Belief and Doubt. *Policy Analysis*, 3(2), 239-271.

Rihoux, Benoit y Ragin, Charles C. (2008). *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques* (Vol. 51). Sage Publications, Inc. <https://us.sagepub.com/en-us/nam/configurational-comparative-methods/book230411>

Schneider, Carsten y Wagemann, Claudius (2012). *Set-Theoretic Methods for the Social Sciences*. Cambridge University Press. <https://www.cambridge.org/core/books/settheoretic-methods-for-the-social-ciencies/236C162386C1188966FE269D625CA289>

Shulha, Lyn M. y Cousins, Bradley (1997). Evaluation Use: Theory, Research, and Practice Since 1986. *Evaluation Practice*, 18(3), 195-208.

Viñas, Verónica (2009). The European Union's Drive towards Public Policy Evaluation: The Case of Spain. *Evaluation*, 15(4), 459-472. <https://doi.org/10.1177/1356389009341900>

Webber, David J. (1984). Political conditions motivating legislators' use of information. *Policy Studies Review*, 4, 110-118. <https://doi.org/10.1111/J.1541-1338.1984.TB00165.X>

Weiss, Carol (1979). The many meanings of research utilization. *Public Administration Review*, 426-431.

Weiss, Carol (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*.
<https://journals.sagepub.com/doi/abs/10.1177/109821409801900103>

Weiss, Carol (1988). If Program Decisions Hinged Only on Information: A Response to Patton. *Evaluation Practice*, 9(3), 15-28. <https://doi.org/10.1177/109821408800900302>

Weiss, Carol, Murphy-Graham, Erin y Birkeland, Sarah (2005). An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation - AM J EVAL*, 26, 12-30. <https://doi.org/10.1177/1098214004273337>

Wilkins, Peter (2012). Cultures of Ombudsman and Audit Institutions: Effects on Their Evaluative Activities. En *Evaluation Cultures: Sense-Making in Complex Times*. Transaction Publishers.

Williams, Kevin, Laats, Bastiaan D. y Stern, Elliot (2002). *The Use of Evaluation in the Commission Services Final Report*. European Commission, DG Budget.

PREPRINT