

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Responder a la complejidad:

Contribuciones de la evaluación y el diseño colaborativo a la innovación de políticas públicas

Carla Cordoncillo Acosta

**DOCTORATS
INDUSTRIALS**

RECERCA ESTRATÈGICA · COL·LABORATIVA · APLICADA

ivàlua 

Institut Català d'Avaluació
de Polítiques Públiques



Responder a la complejidad:

**Contribuciones de la evaluación y el diseño
colaborativo a la innovación de políticas públicas**

Carla Cordoncillo Acosta

Junio de 2025

Tesis doctoral

Mención: Doctorado Industrial

Directora: Raquel Gallego Calderón

PhD in Politics, Policies and International Relations

Institut de Govern i Polítiques Públiques

Facultat de Ciències Polítiques i de Sociologia

Universitat Autònoma de Barcelona

A Max, por haber cambiado mi concepción del tiempo.

Índice

Índice	
Resumen de la tesis	1
Resum de la tesi	2
Dissertation abstract	3
Agradecimientos	4
1. Introducción	5
1.1 Innovar en el diseño de políticas públicas	5
1.2 Objetivos, pregunta de investigación y estructura de la tesis	9
1.3 Contexto de la investigación	12
1.4 Metodología	13
1.5 Referencias.....	16
2. Mechanisms for institutionalising evaluation: a scoping review.....	23
2.1 Introduction	23
2.2 Analytical framework	24
2.3 Methodology	28
2.3.1 Identification of relevant studies	28
2.3.2 Selection and coding of studies	29
2.4 Results.....	30
2.4.1 Characteristics of the studies	30
2.4.2 Identification and categorisation of institutionalisation strategies in political system	31
2.5 Discussion	39
2.6 Conclusion	43
2.7 References	44
2.8 Appendix	50
Appendix 1: Articles included in the scoping review.....	50
3. El uso de la evaluación en un contexto de institucionalización débil: una aproximación desde el <i>Qualitative Comparative Analysis</i>.....	53
3.1 Introducción	53
3.2 Marco teórico y analítico: factores que influyen en el uso de las evaluaciones en contextos de institucionalización avanzada.....	54
3.3 La institucionalización de la evaluación en España	58
3.4 Casos de estudio	59

3.5 Metodología	60
3.5.1 Definición de las condiciones	60
3.5.2 Calibrage.....	62
3.6 Resultados	62
3.6.1 Análisis de la <i>truth table</i>	63
3.7 Discusión	66
3.8 Conclusión	69
3.9 Bibliografía	71
4. Fostering innovation through collaboration: A comparison of collaborative approaches to policy design.....	76
4.1 Introduction	76
4.2 Dealing with complexity from a collaborative approach	77
4.2.1 Empirical evidence	79
4.2.2 Research question and hypotheses	80
4.3 Case selection and data operationalisation	82
4.3 Method.....	85
4.4 Results.....	86
4.5 Discussion	89
4.6 Conclusion	93
4.7 References	94
4.8 Appendix	100
Appendix 1: Score tables measuring innovation (Torfing et al., 2020).....	100
Appendix 2: Distribution of the dependent variable before converting it into a dummy	101
Appendix 3: Regression analysis – innovation as a continuous variable (Table 2A)	101
5. Conclusiones.....	103
5.1 Principales resultados	103
5.2 Relevancia de la tesis	106
5.3 Limitaciones de la investigación	108
5.4 Futuras líneas de investigación	109
5.5 Referencias.....	110

Resumen de la tesis

Esta tesis explora cómo la evaluación y el diseño colaborativo pueden contribuir a la innovación en políticas públicas y dar respuesta a la complejidad y multidimensionalidad de los llamados “problemas malditos”. A través de tres artículos —dos centrados en la evaluación y uno en los enfoques colaborativos— se analiza cómo estas dos estrategias —que tienen orígenes paralelos, pero motivaciones inicialmente contrapuestas— pueden contribuir a mejorar la conceptualización de los problemas y, sobre todo, a plantear mejores soluciones de política pública. Los resultados muestran que las políticas diseñadas de forma colaborativa —independientemente de los actores involucrados— tienen mayor probabilidad de resultar innovadoras que aquellas diseñadas de forma tradicional, por parte de una sola administración. En cuanto a la evaluación, su contribución a la innovación es más indirecta y depende de su integración en procesos de aprendizaje institucional. Responder a la complejidad requiere entender las políticas públicas como hipótesis que deben ser testeadas en la práctica mediante evaluaciones rigurosas que permitan generar inteligencia colectiva. Una aproximación que, lejos de contraponer la evidencia y la evaluación a los enfoques colaborativos, los acerca y apuesta por su sinergia como respuesta a la complejidad.

Resum de la tesi

Aquesta tesi explora com l'avaluació i el disseny col·laboratiu poden contribuir a innovar en política pública i donar resposta a la complexitat i la multidimensionalitat dels anomenats "problemes maleïts". A través de tres articles —dos centrats en l'avaluació i un en els enfocaments col·laboratius— s'analitza com aquestes dues estratègies —que tenen orígens paral·lels però motivacions inicialment contraposades— poden contribuir a millorar la conceptualització dels problemes i, sobretot, a plantejar millors solucions de política pública. Els resultats mostren que les polítiques dissenyades de manera col·laborativa —independentment dels agents implicats— tenen major probabilitat de resultar innovadores que aquelles dissenyades de forma tradicional, per part d'una sola administració. Pel que fa a l'avaluació, la seva contribució a la innovació és més indirecta i depèn de la seva integració en processos d'aprenentatge institucional. Donar resposta a la complexitat requereix entendre les polítiques públiques com a hipòtesis que han de ser testades en la pràctica mitjançant avaluacions rigoroses que permetin generar intel·ligència col·lectiva. Un enfocament que, lluny de contraposar l'evidència i l'avaluació als enfocaments col·laboratius, els apropa i aposta per la seva sinergia com a resposta a la complexitat.

Dissertation abstract

This thesis explores how evaluation and collaborative design can contribute to public policy innovation and help address the complexity and multidimensionality of so-called wicked problems. Through three articles—two focused on evaluation and one on collaborative approaches—it examines how these two strategies, which stem from parallel origins but initially contrasting motivations, can enhance both the conceptualization of problems and, more importantly, the development of more effective policy solutions. The findings show that collaboratively designed policies—regardless of the actors involved—are more likely to lead to innovation than those developed by a single public administration. In contrast, the contribution of evaluation to innovation is more indirect and depends on its integration into institutional learning processes. Addressing complexity requires understanding public policies as hypotheses to be tested in practice through rigorous evaluations that foster the development of collective intelligence. This approach, rather than opposing evidence and evaluation to collaborative methods, seeks to bring them together and promote their synergy as a response to complexity.

Agradecimientos

Esta tesis no hubiera sido posible sin el apoyo de muchísimas personas; razón por la cual quiero darles mis más sinceras y profundas gracias.

A todo el equipo de Ivàlua por acompañarme a lo largo de este viaje. Y especialmente a Marc Balaguer, por darme la oportunidad de hacer un doctorado industrial y por contagiarme su pasión por la institucionalización y el uso de la evaluación. A Mireia Borrell, por tenderme una mano cuando no me veía capaz de seguir sola y por convertir tantas tardes de trabajo en el *Sandwichez* en el inicio de una amistad. A Erika Pérez, por confiar en mí y proporcionarme herramientas para conciliar trabajo, investigación y cuidados. A Andreu Arenas, por su orientación metodológica en un momento de cansancio después de muchas revisiones. A Ariadna Fitó por apoyarme en la recta final. A Júlia de Quintana y Laura López por escucharme, apoyarme y darme ánimo. Y a Anna Segura por encajar incontables cervezas post-pádel cooptadas por el “tema tesis” siempre con interés e ilusión.

A todo el equipo del IGOP, y muy especialmente a mi directora, Raquel Gallego, que no solamente me ha acompañado en cada uno de los pasos y decisiones tomados a lo largo de estos cinco años, sino que me ha ayudado a creer en mí y en mis habilidades como investigadora. Si bien nos hemos visto poco físicamente, siempre ha estado cerca y dispuesta a compartir tanto las buenas noticias como las más difíciles de encajar. ¡Y no hay nadie más ágil haciendo gestiones! También quiero agradecer al resto de compañeros y compañeras de doctorado, con quienes me hubiera gustado compartir más tiempo, por acogerme siempre como si nos viéramos todos los días; gracias David Palomera, Llorenç Soler, Francisco Ferraioli, Maria Mónica Salazar, Adriana Offredi, Matilde Cittadini y Pedroma Bellón. Gracias a Roser Girós, por compartir los retos diarios de llevar a cabo un doctorado industrial. A Armando Buelbas, por compartir alegrías y penas, siempre con una sonrisa. Y, sobre todo, a Yuni Salazar, por hacer de guía: por su cercanía y sus consejos, y por estar siempre dispuesta a tender una mano.

Esta tesis tampoco hubiera sido posible sin todas las amigas y amigos que me acompañaron en este viaje; que se interesaron por mis avances y me dieron fuerza cuando estaba cansada o pensaba que no podía con todo. Gracias Sophia Castellanos y Pipe Pérez por compartir dudas, avances e ilusiones. Gracias, Marcos Perales, Sabrina Rossetti, Alicia Ferreira, Juanita Ardila y Júlia Altés por preguntar siempre. Gracias Marc Folch por ayudarme cuando no sabía ni por dónde empezar y acompañarme en cada paso. Y, sobre todo, gracias Marta Pérez, Xènia Vergès, Carla Longares y Raquel Batet por sostenerme siempre.

Sin embargo, solo mi familia es plenamente consciente de lo que me ha supuesto escribir esta tesis. No habría sido posible sin el ejemplo de mi padre: tu constancia y ambición, papa, han sido un espejo en el que mirarme. Ni tampoco sin el apoyo incondicional de mi madre. Gracias, mama, por incontables domingos por la tarde con Max para que pudiéramos estudiar; sin tu apoyo y tus cuidados no hubiera podido seguir, ni mucho menos terminar. Y por último, gracias Juanjo por ser mi compañero de viaje. Por compartir infinitas tardes de estudio y noches con el cerebro fundido, cuando estábamos motivados y cuando estábamos agotados. Por celebrar cada logro como si fuera tuyo. Por entender el valor, pero también la relatividad, de esta tesis. Y por hacerme creer, desde el primer día, que podía lograrlo. Somos un gran equipo.

1. Introducción

1.1 Innovar en el diseño de políticas públicas

El contexto actual está marcado por problemas sociales que se caracterizan por ser cambiantes, interdependientes, multi-causales, intrínsecamente políticos y, por lo tanto, difíciles de definir e imposibles de resolver de forma definitiva; lo cual les otorga el calificativo de problemas “malditos” (los famosos *wicked problems*) (Geyer & Rihani, 2010; Rittel & Webber, 1973; Wagenaar, 2007). Los ejemplos más habituales son el cambio climático, los movimientos migratorios globales, la persistencia de la pobreza o el envejecimiento de la población, entre muchos otros.

Si bien el concepto de “problemas malditos” tiene ya más de cincuenta años (Rittel & Webber, 1973), la realidad es que a día de hoy **los gobiernos continúan sin saber cómo responder a esta complejidad**. Aunque los problemas actuales son claramente multinivel y multi e intersectoriales, las administraciones públicas continúan estructuradas, no solamente de forma claramente jerárquica, sino también en silos (Crosby et al., 2016), a partir de Ministerios o Departamentos sectoriales; lo cual dificulta enormemente, no solo la incorporación de cualquier perspectiva transversal, sino también el planteamiento de soluciones integrales (Agranoff, Robert, 2003; Weber & Khademian, 2008).

Además, la literatura apunta que a menudo, las decisiones “incompetentes o a corto plazo de los gobiernos resultan no solamente en intervenciones sin efecto sino en un empeoramiento de estos problemas” (Head, 2022, citando a King y Crewe, 2014). En algunos casos, una mala gestión puede llevar incluso a convertir en “malditos” problemas que de entrada no lo eran: lo que se ha llamado problemas “malditos por diseño” (*wicked by design*), fundamentalmente porque los líderes políticos utilizan problemas bien definidos como “sustitutos para debatir problemas mayores y más controvertidos” (Head, 2022 citando a Nie, 2003).

A la difícil gestión de esta complejidad se suma una mayor demanda de la ciudadanía de participar e **involucrarse en los procesos de decisión**, que conlleva también un aumento de sus expectativas respecto a las acciones gubernamentales (Bentzen et al., 2020). Los ciudadanos y ciudadanas esperan no solamente incidir en la agenda política, sino también proponer y discutir opciones de política pública (OECD, 2001).

Ante este escenario, la innovación de políticas públicas no solamente se ha vuelto indispensable (Albury, 2011; Bason, 2017; Bourgon, 2011; Mulgan, 2014), sino que se ha convertido en uno de los principales retos que enfrentan las democracias contemporáneas (Head, 2022). Según Torfing (2016), esta innovación implica introducir nuevos objetivos, instrumentos y herramientas que rompan con las prácticas convencionales, con el propósito de encontrar soluciones de política pública más efectivas. Las exigencias de mayor participación y la complejidad e interconexión de los problemas requieren además nuevos enfoques. Sin embargo, contra más difíciles de gestionar son los problemas, menos probable es que los gobiernos estén dispuestos a asumir riesgos (Bhatta, 2003). Por esta razón resulta especialmente relevante entender qué factores pueden propiciar esa innovación de políticas públicas.

Si la innovación debe permear el *policymaking*, es fundamental que la disrupción ocurra desde el inicio del ciclo de la política pública, específicamente en la fase de diseño. Este es el momento tanto para analizar y comprender los problemas —"cuando los actores intentan persuadir a los decisores de política pública sobre la naturaleza y el significado de los temas en discusión" (Head, 2022a:9)— como para plantear posibles soluciones. Howlett (2014) entiende esta fase del **diseño de la política pública** como el momento en el cual los *policymakers* aplican los conocimientos de los que disponen sobre los *policy means*. Sin embargo, en muchos casos, "la forma como los diferentes actores y los decisores de política pública definen un problema está fuertemente correlacionada con sus preferencias por herramientas o soluciones específicas" (Head, 2022a:10). En otras palabras, la forma como se interpretan y definen los problemas está directamente relacionada con las soluciones que se implementan como respuesta. Por eso es sumamente importante que durante ese proceso de definición (*framing*) y de análisis de posibles soluciones los *policymakers* y los *decisión-makers* dispongan de información rigurosa, relevante y útil, que les permita plantear soluciones efectivas e innovadoras.

Ante esta necesidad, **cobran fuerza dos tipos de estrategias que**, a pesar de tener distinto origen y naturaleza, **comparten el objetivo de proponer mejoras en el diseño de las políticas públicas**. Se trata, por un lado, del **Movimiento de Políticas Basadas en Evidencia** (EBPM, según sus siglas en inglés), que defiende la importancia de que las políticas públicas se diseñen en base al conocimiento científico disponible, y, por lo tanto, a información rigurosa, producida siguiendo unos estándares académicos. El Oxford English Dictionary define la palabra "evidencia" como "el cuerpo disponible de hechos o informaciones que permiten indicar si una creencia o proposición es cierta o válida" (de Marchi et al., 2016:3). Así, el EBPM se ha definido como un enfoque que "ayuda a tomar decisiones informadas sobre políticas públicas, programas y proyectos, poniendo a disposición la mejor evidencia científica disponible" (de Marchi et al., 2016, citando a Davies, 1999). Un enfoque que propone dejar atrás los procesos de toma de decisión basados en la *opinión* para priorizar aquellos basados en la *evidencia* (Davies, 1999).

Este movimiento, que se popularizó con el Gobierno de Toni Blair en el Reino Unido (bajo el eslogan "*what matters is what works*") (Banks, 2009), se extendió y asentó en todas las democracias occidentales de la mano de las premisas de eficacia y eficiencia que defendió —y en buena medida continua defendiendo a día de hoy— el *New Public Management* (Pattyn et al., 2017): el paradigma de gestión pública que se presentó como alternativa a la "administración pública tradicional" prácticamente en todo el mundo entre los años 1980 y 1990 (Bryson & Crosby, 2014). Planteado como una respuesta a la ineficacia gubernamental (en un contexto de crisis económica) y epistemológicamente fundamentado en el positivismo, el *New Public Management* se basó en la racionalidad económica para defender desde la adopción de instrumentos de gestión privada en el sector público hasta la privatización de la gestión de los servicios públicos (Bryson & Crosby, 2014).

En ese contexto, los defensores y defensoras de las políticas basadas en evidencias propusieron aplicar los acercamientos y los métodos de resolución de problemas propios de los ámbitos de la ingeniería y de la salud, a los problemas sociales y económicos (Head, 2022). La idea principal, detrás de esa **concepción racional del policymaking** es que el conocimiento científico facilita la comprensión de tendencias, el

establecimiento de relaciones causales y la medición de riesgos y de potenciales impactos de intervenciones concretas (Head, 2022). Si entendemos los mecanismos que determinan la consecución de resultados deseados concretos, podremos diseñar políticas que funcionen mejor (Pawson & Tilley, 1997).

Ello explica que el auge de las políticas basadas en evidencias haya dirigido la atención hacia un tipo particular de evidencia: la **evaluación de políticas públicas** (Pattyn et al., 2017), definida como “el análisis científico de una política pública o una parte de esta, cuyo objetivo es determinar el mérito o valor de esa política en base a una serie de criterios” (Pattyn, 2014:44). La evaluación es una pieza fundamental de la “caja de herramientas” que construyeron las administraciones públicas en tiempo de crisis económica para la racionalización del gasto público (Olejniczak et al., 2016:92). Hasta el punto de que en los últimos 20 años —y en prácticamente la totalidad del planeta— la evaluación de políticas públicas se ha convertido en un requisito indispensable para el buen funcionamiento de las administraciones públicas (Dahler-Larsen, 2012).

La contribución de la evaluación a las políticas públicas tiene tres formas: (1) establecer qué soluciones funcionan, (2) explicar por qué las soluciones funcionan (o no) —y por lo tanto identificar y entender los mecanismos causales que determinan ese bien o mal funcionamiento— y (3) transferir los hallazgos de la evaluación (la evidencia) a las políticas públicas (Olejniczak et al., 2016). Siguiendo esta lógica, el enfoque de políticas basadas en evidencias busca que el ciclo de gestión de las políticas públicas (*policy cycle*) empiece con un debate informado en evidencia sobre los problemas de política pública y las posibles soluciones y termine con una evaluación rigurosa de sus soluciones (Cairney, 2016:14), que a su vez sirva para informar un nuevo debate.

De esta forma, la evaluación se convierte en un elemento integral y transversal del proceso de formulación e implementación de políticas públicas (*policy process*) (Howlett, 2019; Olejniczak et al., 2016). Adicionalmente, en la medida en que promueve el aprendizaje y, por lo tanto, la mejora de la utilidad de las intervenciones públicas (Mark et al., 2000), **contribuye también a ofrecer mejores soluciones a los problemas sociales complejos**. De hecho, “la finalidad última de la evaluación es “la mejora de la sociedad” a través del aprendizaje vinculado a las políticas públicas (*policy learning*)” (Olejniczak et al., 2016: 92, citando a Henry y Mark, 2003). Y con el tiempo han ido emergiendo enfoques y métodos de evaluación específicamente orientados a promover la innovación para adaptarse a realidades dinámicas y contextos complejos. Por ejemplo, la llamada evaluación realista asume que los problemas complejos raramente tienen una solución única y que, en cualquier caso, el funcionamiento de ésta siempre dependerá del contexto; de esta forma, se focaliza en la interacción entre los mecanismos causales y el contexto específicos para entender cómo y por qué funcionan (o no) las políticas. De forma similar, Patton (2011:7) desarrolla la idea de *developmental evaluation* justamente para hacer frente a situaciones de incertidumbre y “dar sentido a aquello que ocurre en condiciones de complejidad, documentando e interpretando dinámicas, interacciones e interdependencias que se dan a medida que se desarrollan las políticas innovadoras”.

Por otro lado, ya en los años 80, Mel Webber (uno de los creadores del concepto “problemas malditos”) planteaba la “descentralización” de los procesos de toma de decisión —haciéndolos más plurales y receptivos— como una vía alternativa a la

racionalidad científica y tecnocrática, pero también orientada a producir mejores resultados de política pública y adaptarse al cambio (Head, 2022a, citando a Webber, 1983). Este es un acercamiento consistente con las propuestas que desarrollaron durante los años 1980 y 1990 los defensores y defensoras de un *policymaking* más participativo (Fischer, 1993; Fischer & Forester, 1993; Healey, 1997; Innes & Booher, 1999) y que, no solamente constituyó una de las principales críticas al paradigma del *New Public Management*, sino que está en el origen del nuevo paradigma de gestión pública —emergente desde los 2000— denominado *New Public Governance* (Bingham et al., 2005; Bryson & Crosby, 2014; Osborne, 2006; Torfing et al., 2020). Alrededor de este cambio de paradigma han emergido numerosos conceptos —como por ejemplo, *collaborative public management* (Agranoff & McGuire, 2003), *joined-up government* (Carey & Crammond, 2015), *networked government* (Goldsmith & Eggers, 2004), *networked governance* (Stoker, 2006) o *collaborative governance* (Emerson et al., 2012; Paquet, 2009)—, que a su vez se han dotado de herramientas que han adquirido tanta relevancia que han generado su propia línea de investigación académica; es el caso de los conceptos *co-production* (Alford, 2016; Brandsen, 2018; Sorrentino et al., 2018), *co-creation* (Ansell & Torfing, 2021; Torfing et al., 2016, 2021) o *collaborative innovation* (Ansell & Torfing, 2014; Nambisan, 2008; Torfing, 2016).

Aunque cada una de estas propuestas tiene sus particularidades, lo que resulta interesante en el marco de esta investigación es que, quienes abogan por estos **enfoques colaborativos** con relación al *policymaking*, argumentan que entre sus bondades se encuentra su **potencial para proporcionar respuestas y soluciones innovadoras a los problemas sociales**. En la medida en qué la colaboración permite la interacción de diferentes perspectivas y deriva en la generación de sinergias —entre los distintos actores implicados en el diseño de las políticas—, posibilita la generación y materialización de unas ideas que esos mismos actores no podrían alcanzar individualmente (Bason, 2010; Straus, 2002; Torfing, 2016). Más allá de la fase de diseño de las políticas o soluciones —cuando emergen esas ideas innovadoras—, la colaboración también facilita la implementación de políticas en contextos complejos y de constante cambio, pues, a partir de la construcción de plataformas de diálogo, permite anticipar y ajustar debilidades de forma rápida y sencilla, y establecer procesos adaptativos (Ansell & Gash, 2007, 2018).

Como se ha comentado, esta idea cobra tanta fuerza en la academia que a principios de los 2000 emerge un ámbito de investigación específico alrededor del concepto de “**innovación colaborativa**” (*collaborative innovation*). Nambisan (2008:11) la define como un “enfoque colaborativo de la innovación y la resolución de problemas en el sector público, que se basa en aprovechar los recursos y la creatividad de redes y comunidades externas (incluyendo redes de ciudadanos, así como redes de organizaciones sin fines de lucro y corporaciones privadas) para amplificar o mejorar la velocidad de la innovación, así como la gama y la calidad de los resultados de la innovación”.

Es decir, la innovación colaborativa se construye sobre la idea del “gobierno en red” (*networked government*) (Bommert, 2010) —posteriormente se relaciona también con la gobernanza en red (*networked governance*) (Crosby et al., 2016; Sørensen & Torfing, 2011; Torfing, 2016)—, y esta gestión en red (*networked management*), descentralizada, se relaciona positivamente con la capacidad de los gobiernos de

encontrar soluciones a los problemas complejos (Agranoff, 2007). Los problemas actuales no pueden descomponerse de forma que una única organización pública —ni siquiera un gobierno— asuma su resolución (Bason, 2010; Crosby et al., 2016; Doz & Koskonen, 2014), sino que por su complejidad, especialmente los “problemas malditos”, **exigen soluciones multi-actor** (Ansell & Torfing, 2014; Bason, 2010; Bingham et al., 2005; Doz & Koskonen, 2014; Emerson et al., 2012; Hartley, 2005; Osborne, 2006). Ante esa realidad, “los *public managers* deben convertirse en orquestadores de las interacciones de la red y promotores del aprendizaje mutuo: actuando como convocadores, administradores y catalizadores de la innovación colaborativa” (Crosby et al., 2016:2).

En definitiva, con motivaciones y orígenes muy distintos, desde principios de los 2000, el diseño colaborativo y la evaluación de políticas públicas se han consolidado —tanto a nivel académico como entre la mayoría de las administraciones públicas— como vías para innovar en política pública y encontrar mejores soluciones a los problemas sociales complejos que caracterizan nuestra sociedad.

1.2 Objetivos, pregunta de investigación y estructura de la tesis

Ante esta constatación, el objetivo de esta tesis es **investigar la contribución, tanto de la evaluación como del diseño colaborativo de políticas públicas, a la innovación de políticas públicas**. El análisis de la relación de ambos instrumentos —por separado— con la innovación de políticas públicas no es una novedad, aunque la evidencia existente cuenta con limitaciones que esta tesis pretende minimizar. Por un lado, existen limitaciones metodológicas en el caso del diseño colaborativo —pues la evidencia empírica cuantitativa que sustenta la relación es tremendamente escasa. Por otro lado, existen limitaciones epistemológicas y conceptuales, en el caso de la evaluación, cuya relación directa con la innovación de políticas públicas está muy poco explorada, aunque bajo una visión racionalista todavía muy extendida tiende a asumirse que contribuye a la mejora continua y por lo tanto a una cierta innovación incremental.

Sin embargo, lo que sí resulta novedoso es **abordar la contribución de ambos instrumentos conjuntamente**. Como se ha expuesto en el apartado anterior, el Movimiento de Políticas Basadas en Evidencias y los enfoques colaborativos sobre la política pública se desarrollan de forma paralela —y por lo tanto muy poco interconectada. De hecho, no es casualidad que ambos encuentren sus orígenes en los años 70s, coincidiendo con la publicación de Rittel y Webber (1973) del artículo “*Dilemmas in a general theory of planning*”, en el cual teorizan por primera vez alrededor del concepto de *wicked problems*. Fue en ese momento de crisis del paradigma de administración pública tradicional (Dunleavy & Hood, 1994; Hood, 1995) o de gestión pública burocrática —evidenciada por grandes movilizaciones y protestas a nivel social, que pusieron en entredicho la efectividad de las políticas sociales y del keynesianismo económico y se tradujeron en una decaída importante de la confianza en las instituciones (Osborne, 2006)— cuando Rittel y Webber conceptualizaron por primera vez las limitaciones del racionalismo científico que caracterizaba la burocracia tradicional para resolver la complejidad de los problemas sociales de ese momento (Crowley & Head, 2017).

Como respuesta, por un lado emergió con fuerza el Movimiento de Políticas Basadas en Evidencias —pregonando un enfoque más bien empiricista, según el cual la política pública se puede mejorar analizando científicamente qué funciona—, que además coincidió en el tiempo con el reconocimiento de la comunidad académica del llamado “*utilisation gap*”, es decir, del hecho de que el conocimiento producido a nivel científico y académico no se estaba incorporando en los procesos de toma de decisión política (C. H. Weiss, 1988). Por el otro, sobre todo a nivel académico, se empezaron a forjar una serie de ideas y métodos —como el constructivismo o los métodos argumentativos y participativos (promovidos por referentes como Paulo Freire, Orlando Fals Borda o Peter Reason)— que años después se materializarían en los enfoques colaborativos sobre la política pública. Pero ambos defendieron ofrecer mejores soluciones de política pública y, a medida que el concepto de *wicked problems* fue ganando relevancia a nivel académico y que el tema de la complejidad de los problemas sociales se convirtió en una de las principales preocupaciones de las administraciones públicas, ambos movimientos desarrollaron estrategias para intentar dar respuesta a esa complejidad. Por esta razón esta tesis propone poner el énfasis en la intención de ambas estrategias de ofrecer mejores soluciones de política pública y analizar de forma conjunta su contribución a la innovación de políticas públicas. Más concretamente, la pregunta de investigación que esta tesis busca responder es: *¿Cómo pueden la evaluación y el diseño colaborativo contribuir a la innovación de políticas públicas? ¿De qué maneras pueden ayudar a los policymakers a gestionar mejor la complejidad de los problemas sociales actuales?*

Para dar respuesta a estas preguntas, la tesis se estructura en tres capítulos, que constituyen tres artículos independientes —dos de los cuales publicados en el momento del depósito y el tercero en proceso de publicación—, más esta introducción y una conclusión. Los dos primeros artículos se centran en la contribución de la evaluación y el tercero en la contribución de los enfoques colaborativos al diseño de políticas públicas.

Puede parecer obvio, pero para que la evaluación pueda contribuir a mejorar las políticas públicas de las tres formas que se apuntaban anteriormente (identificando qué funciona, por qué funciona y traspasando ese conocimiento a las políticas públicas en funcionamiento) **primero hay que evaluar**. Aunque se ha comentado que la evaluación es hoy en día una práctica extendida (Dahler-Larsen, 2012) y diversos organismos internacionales promueven su desarrollo (OCDE, Unión Europea, Banco Mundial, entre otros), su institucionalización —entendida como el despliegue de mecanismos dirigidos a lograr que sea una práctica recurrente (Pattyn et al., 2017)— sigue siendo más bien débil o, en cualquier caso, reciente en la mayor parte de países (Furubo et al., 2002; Jacob et al., 2015; Stockmann et al., 2020b, 2022, 2023).

De hecho, los principales marcos analíticos utilizados hasta la fecha para estudiar la institucionalización de la evaluación (Furubo et al., 2002; Jacob et al., 2015; Meyer et al., 2020; OECD, 2020; Varone & Jacob, 2004) han sido diseñados precisamente para medir —a partir de índices— y comparar el nivel de institucionalización de distintos países. Sin embargo, una comprensión profunda de las estrategias que los países pueden desarrollar para hacer de la evaluación una práctica sistemática y recurrente requiere ir más allá de los indicadores e identificar, también, cuáles son los mecanismos y los arreglos institucionales más comúnmente utilizados. Si bien las recientes compilaciones

de Stockmann y Meyer (Stockmann et al., 2020, 2022, 2023) —que son sin duda el ejercicio más exhaustivo realizado sobre el tema— recogen buena parte de esta información, ni su marco analítico permite llegar a este nivel de detalle, ni el tipo de comparaciones que realizan al final de cada volumen son sistemáticas. El **primer artículo** de esta tesis (Capítulo 2) pretende llenar este vacío identificando y categorizando —a partir de una *scoping review*— los mecanismos de institucionalización de la evaluación existentes y reflexionando sobre qué factores deben tener en consideración los *public managers* a la hora de seleccionarlos para fortalecer sus respectivos sistemas de evaluación.

Una vez creadas las condiciones para que las administraciones públicas hagan de la evaluación una práctica sistemática y recurrente es necesario garantizar que estas evaluaciones generan información útil y derivan en aprendizajes que son tenidos en cuenta para mejorar las políticas públicas. Si bien **informar los procesos de decisión** es la *raison d'être* de la evaluación, tanto los estudios comparados sobre su institucionalización (Stockmann et al., 2020a; Stockmann & Meyer, 2022a, 2022b) como la evidencia empírica disponible sobre su uso instrumental (Hornby & Perera, 2002; Pattyn & Bouterse, 2020) apuntan que el llamado *utilisation gap* continúa siendo el principal reto que enfrentan los sistemas de evaluación prácticamente en todos los países, pero especialmente en aquellos con menor tradición de evaluación.

Por esta razón, el **segundo artículo** de esta tesis (Capítulo 3) busca identificar *qué factores explican el uso instrumental de las evaluaciones en contextos con una institucionalización débil*. A diferencia de la mayoría de los estudios previos, elaborados en países con una cultura de evaluación más madura, este artículo complementa la evidencia existente al analizar entornos donde la evaluación está menos arraigada. Los resultados muestran que, en estos contextos, adquieren mayor relevancia aquellas condiciones diseñadas específicamente para compensar la falta de estructuras formales. Así, al identificar qué factores facilitan que los resultados de las evaluaciones sean considerados en los procesos de toma de decisiones, es posible orientar a los *policymakers* para potenciar el carácter transformador de la evaluación, especialmente en contextos donde su institucionalización es aún incipiente.

El **tercer artículo** (Capítulo 4 de la tesis) se centra en la contribución de los enfoques colaborativos a la innovación en el diseño de políticas públicas. Más concretamente, el artículo busca complementar la evidencia empírica disponible —de carácter exclusivamente cualitativo— que relaciona la colaboración entre distintos actores en el diseño de las políticas públicas con la innovación, con evidencia cuantitativa sobre *qué tipo de enfoques colaborativos tienen mayor potencial de resultar en políticas innovadoras*. Así, el análisis permite, no solamente validar la relación entre colaboración e innovación de políticas públicas, sino también identificar qué actores (administraciones públicas, entidades del tercer sector, empresas privadas, asociaciones de vecinos/as y ciudadanía no organizada) y combinaciones de actores tienen mayor probabilidad de desarrollar políticas innovadoras.

Finalmente, en el **Capítulo 5** se discuten los resultados de la tesis, se ponen de manifiesto tanto su contribución como sus limitaciones y se identifican posibles líneas de investigación a partir de las cuales continuar profundizando sobre la relación entre evaluación, diseño colaborativo e innovación de políticas públicas.

1.3 Contexto de la investigación

Esta investigación se realiza en el marco de una **Beca de Doctorado Industrial (2020-2024)** de la Agencia de Gestión y Ayudas Universitarias y de Investigación (AGAUR), que se concreta en un convenio de colaboración entre la Universitat Autònoma de Barcelona (UAB) —y más concretamente el Instituto de Gobierno y Políticas Públicas (IGOP)— y el Institut Català d’Avaluació de Polítiques Públiques (Ivàlua), donde trabajo como analista desde el mes de enero de 2020, antes de iniciar la tesis doctoral en septiembre de ese mismo año.

Los doctorados industriales tienen por objetivo combinar la investigación académica con la aplicación profesional y práctica en un entorno empresarial o una administración pública y, por lo tanto, en el ámbito específico de las políticas públicas, suponen una oportunidad única para acortar la distancia existente entre las llamadas “dos comunidades” (Caplan, 1979), es decir, entre la academia y las administraciones públicas. De acuerdo con Mead (2015), el conocimiento científico no se utiliza para informar las políticas públicas porque la academia no produce información relevante para los *policymakers*. Más concretamente, el autor considera que: (1) la investigación académica no se alinea con los discursos políticos del momento, (2) la academia no ofrece estrategias políticas que ayuden a quienes toman las decisiones a implementar propuestas de política pública, (3) además escribe en un lenguaje que los *policymakers* no entienden y (4) los investigadores e investigadoras tiene pocos incentivos profesionales para relacionarse activamente con los gobiernos, porque lo único que se valora para su proyección profesional es publicar en revistas académicas (Newman & Head, 2015).

En ese sentido, **Ivàlua ofrece una posición privilegiada para entender cómo funciona la gestión de las políticas públicas desde dentro de la administración, sin necesidad de renunciar al lenguaje, la mentalidad ni el propósito de la academia.** De hecho, esas mismas diferencias (entre quienes producen el conocimiento científico y quienes deben utilizarlo para mejorar las políticas públicas) han sido identificadas como una de las causas del *utilisation gap* (Olejniczak et al., 2016); es decir, como uno de los factores que explican que los resultados de las evaluaciones raramente sean tenidos en cuenta para informar los procesos de toma de decisión.

Así, estudiar el potencial de la evaluación para innovar en el diseño de políticas públicas desde una institución dedicada específicamente a evaluar y promover la cultura de la evaluación no solamente tenía todo el sentido —en términos de misión institucional—, sino que ofrecía una doble oportunidad en términos prácticos: por un lado, la posibilidad de contrastar empíricamente buena parte de los debates y retos que se plantean desde la literatura académica; por otro lado, una cercanía con el día a día del *policymaking* que abría la puerta a contribuir a esa literatura con datos, experiencias y reflexiones prácticas.

Es por esta razón que **cada uno de los artículos recogidos en esta tesis tiene entre sus objetivos la intención de ofrecer soluciones y herramientas concretas a los *policymakers*, en base conocimiento teórico y empírico existente.** El primer artículo plantea una categorización de mecanismos de institucionalización que, si bien tiene también un interés académico, está pensada para llenar un vacío práctico, para facilitar

la toma de decisiones a quienes se encargan de impulsar la institucionalización de la evaluación en las administraciones públicas. Los resultados del segundo artículo deberían servir tanto a evaluadores/as como, sobre todo, a *policymakers* para impulsar procesos de evaluación con mayor potencial de uso instrumental; es decir, para desarrollar evaluaciones que tengan mayor probabilidad de terminar informando el diseño de la política pública. Por último, el tercer artículo, busca explícitamente facilitar a los *policy designers* la decisión sobre quién invitar a participar en el diseño de las políticas públicas. De esta forma, el espíritu del doctorado industrial impregna el contenido y la voluntad de esta tesis de contribuir a acercar esas “dos comunidades”.

Mientras desarrollaba esta tesis doctoral he participado en la evaluación de numerosas políticas públicas. Entre ellas, destacan la primera edición del *Pla de Barris* del Ayuntamiento de Barcelona, las Leyes autonómicas 17/2015 de igualdad efectiva entre mujeres y hombres, y 19/2020 de igualdad de trato y no-discriminación, el programa de cuidados a la infancia *Temps x Cures* (impulsado por la Dirección General de Cuidados, Organización del Tiempo y Equidad en los Trabajos del Departamento de Feminismos de la Generalitat) o los servicios de atención diurna para personas con discapacidad incluidos en la Cartera de servicios sociales (e impulsados por el Departamento de Derechos Sociales). Así mismo, he asesorado a diferentes unidades de la Generalitat de Catalunya en el ámbito de la institucionalización de la evaluación (Dirección General de Cooperación al Desarrollo, Departamento de Justicia, Consorci per la Normalització Lingüística) y he coordinado el diseño y la implementación de dos iniciativas *Qué Funciona* (un *Qué funciona en cuidados de larga duración* con la Diputació de Barcelona, la Taula del Tercer Sector y la Confederació del Tercer Sector, y un *Qué funciona en género y juventud*, con la Agencia Catalana de Juventud y el Instituto Catalán de la Mujer) en el marco de las cuales se han generado diversas revisiones sistemáticas de evidencia. También he participado como docente en numerosos programas de formación sobre evaluación dirigidos tanto a funcionarios (de administraciones locales y autonómicas) como a estudiantes de máster de diferentes universidades, y he diseñado una herramienta —de momento interna, pero que tiene intención de ser escalable más allá de l’Àlvia— de seguimiento del uso de las evaluaciones.

Todas estas experiencias se han nutrido —en mayor o menor medida, y de forma más o menos directa— de los conocimientos adquiridos en el proceso de realización de esta investigación. Al mismo tiempo, esta tesis tendría una forma, un contenido y una intención muy diferente si no se hubiera realizado de la mano de todas esas iniciativas. Si bien desarrollar ambas funciones al mismo tiempo resultó un reto importante, fue mi crecimiento como evaluadora lo que me permitió identificar cuál podía ser mi contribución como investigadora.

1.4 Metodología

Cada uno de los artículos que componen esta tesis busca responder a una pregunta de investigación específica, a partir de una metodología concreta. El primer artículo opta por una **scoping review** para identificar de forma tanto exhaustiva como sistemática los mecanismos de institucionalización de la evaluación existentes. Se trata de un tipo de

revisión sistemática que resulta especialmente adecuado en ámbitos de conocimiento todavía emergentes (Mak & Thomas, 2022) —como es el caso de la institucionalización de la evaluación—, pues es un poco más flexible que las revisiones sistemáticas tradicionales y resulta útil para mapear el conocimiento existente. Así, partiendo de una selección de 12 países con tradiciones de evaluación distintas, la revisión permite, primero, construir un listado de mecanismos de institucionalización de la evaluación, para después organizarlo en categorías que contribuyen a avanzar en la conceptualización de la institucionalización. En definitiva, permite sintetizar la información existente en un formato sencillo que puede resultar especialmente adecuado para los *policymakers*.

Más allá de esta investigación, aprender a realizar una revisión sistemática me ha resultado especialmente útil porque —como he mencionado— como analista de Ivàlua coordino dos proyectos *Qué Funciona*, que tienen por objetivo precisamente sintetizar la evidencia existente en ámbitos de conocimiento específico (cuidados de larga duración y género y juventud) para ponerla en formatos accesibles y acercarla a quienes diseñan e implementan políticas y programas en esos ámbitos de intervención. Si bien las revisiones de evidencia se encargan a investigadores e investigadoras expertas en cada materia, una de mis funciones como coordinadora es garantizar la calidad metodológica de las síntesis. En ese sentido, realizar yo misma una revisión sistemática me ha permitido acompañar mejor a esas investigadoras, lo cual es también positivo para Ivàlua.

El segundo artículo utiliza un ***Qualitative Comparative Analysis*** para explorar las condiciones necesarias y suficientes, así como las configuraciones causales, que determinan el uso instrumental de las evaluaciones. Según su creador, Charles Ragin, el QCA es un enfoque, un método específico y una técnica de análisis de datos al mismo tiempo (Befani, 2016; Varone et al., 2006). Como enfoque, sus defensores señalan que combina las fortalezas de los enfoques cuantitativos y cualitativos, y es especialmente adecuado para estudios con un número intermedio de observaciones (Ragin, 1989, 2014). Como técnica, utiliza álgebra booleana para reducir la complejidad de las relaciones causales a través de un procedimiento llamado minimización booleana, que concluye en una serie de combinaciones que llevan a un resultado positivo y/o negativo (Varone et al., 2006). Además, como método comparativo, "busca un equilibrio entre la complejidad y la generalización" (Khagram et al., 2009:259) y propone un punto intermedio entre la "universalización" de los enfoques macro-cuantitativos y la individualización de los enfoques orientados a casos (Berg-Schlosser & Cronqvist, 2005; Ryan & Smith, 2013).

Si bien ha ido ganando peso en el ámbito de las políticas públicas (Rihoux et al., 2011), se trata de un método todavía poco utilizado que, no solamente permite tender un puente entre las metodologías cuantitativas y las cualitativas (Mello, 2021), sino que, además, al enfocarse en muestras medianas (idealmente entre 8 y 40 casos) cubre un vacío metodológico importante, por lo menos en el ámbito de la evaluación de políticas públicas —en el cual los estudios comparados son poco habituales y prevalecen o bien los estudios de caso o bien los análisis estadísticos que requieren muestras grandes. En ese sentido, la exploración del *Qualitative Comparative Analysis* como parte de esta tesis ha supuesto un valor agregado en términos metodológicos para Ivàlua, que ha

podido incorporar el método a su “caja de herramientas” para la evaluación de políticas públicas y, de hecho, ya lo ha aplicado en una evaluación.

Adicionalmente, resulta un método especialmente adecuado para estudiar el uso instrumental de las evaluaciones pues, no solamente permite considerar factores contextuales, sino que, además, y a diferencia de los métodos estadísticos, permite entender la causalidad de forma compleja. Lo hace teniendo en cuenta dos principios causales: la equifinalidad y la causalidad coyuntural (Rihoux & Ragin, 2008). Tal como se explica en el mismo artículo, 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 compleja de entender la causalidad (o por lo menos más compleja que aquella subyacente a los métodos estadísticos) resulta especialmente adecuada porque la literatura teórica apunta que los usos de la evaluación dependen de múltiples factores (Pattyn & Bouterse, 2020; C. Weiss, 1998).

El tercer artículo utiliza un análisis **de regresiones** para comparar el potencial de innovación de 529 intervenciones de política pública impulsadas por el Ayuntamiento de Barcelona. De esta forma, la investigación complementa la evidencia existente sobre la relación entre la colaboración y la innovación en política pública —eminentemente cualitativa y basada en estudios de caso— con un análisis cuantitativo. El análisis de regresiones permite formular, a partir de la literatura teórica, diversas hipótesis sobre la capacidad de innovar de una serie de agentes que participan en el diseño de las políticas públicas, y contrastarlas empíricamente con una muestra suficientemente grande.

Como investigadora, publicar un artículo basado en un análisis de regresiones me ha obligado a reforzar mis habilidades de análisis de datos cuantitativos, lo cual a su vez me ha resultado especialmente útil para poder coordinar —en el marco de mi trabajo como analista en Ivàlua— evaluaciones mixtas, que utilizan métodos tanto cualitativos como cuantitativos.

Juntos, estos tres tipos de métodos no solo complementan sus respectivas fortalezas y mitigan sus limitaciones, sino que además facilitan una comprensión holística y profunda de cómo se diseñan, implementan y evalúan las políticas públicas. De esta forma, la misma tesis, en su conjunto, constituye una exploración de las aportaciones que cada tipo de método puede realizar para mejorar la comprensión del funcionamiento de las políticas públicas en contextos complejos. Esta exploración es relevante en sí misma, pues —a pesar de esfuerzos concretos anteriormente mencionados, como la evaluación realista y la *developmental evaluation*— “los modelos tradicionales de evaluación se han mostrado ineficaces para captar la complejidad creciente” (Uusikylä et al., 2020:84). “La evaluación tradicional aspira a controlar y predecir, a aportar orden al caos” (Patton, 2011:5).

1.5 Referencias

Agranoff, R. (2007). *Managing within Networks. Adding Value to Public Organizations*. Georgetown University Press. <http://press.georgetown.edu/book/georgetown/managing-within-networks>

Agranoff, R., & McGuire, M. (2003). *Collaborative Public Management: New Strategies for Local Governments*. Georgetown University Press. <http://www.jstor.org/stable/j.ctt2tt2nq>

Agranoff, Robert. (2003). *Leveraging Networks: A Guide for Public Managers Working across Organizations*. New Ways to Manage Series. IBM Endowment for The Business of Government. <http://www.businessofgovernment.org/report/leveraging-networks-guide-public-managers-working-across-organizations>

Albury, D. (2011). Creating the Conditions for Radical Public Service Innovation1. *Australian Journal of Public Administration*, 70(3), 227-235. <https://doi.org/10.1111/j.1467-8500.2011.00727.x>

Alford, J. (2016). Co-Production, Interdependence and Publicness: Extending public service-dominant logic. *Public Management Review*, 18(5), 673-691. <https://doi.org/10.1080/14719037.2015.1111659>

Ansell, C., & Gash, A. (2007). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory*, 18(4), 543-571. <https://doi.org/10.1093/jopart/mum032>

Ansell, C., & Gash, A. (2018). Collaborative Platforms as a Governance Strategy. *Journal of Public Administration Research and Theory*, 28(1), 16-32. <https://doi.org/10.1093/jopart/mux030>

Ansell, C., & Torfing, J. (2014). *Public Innovation through Collaboration and Design* (p. 246). <https://doi.org/10.4324/9780203795958>

Ansell, C., & Torfing, J. (2021). Co-creation: The new kid on the block in public governance. *Policy & Politics*, 49(2), 211-230. <https://doi.org/10.1332/030557321X16115951196045>

Banks, G. (2009). *Evidence-Based Policy Making: What is It? How Do We Get It?* (SSRN Scholarly Paper ID 1616460). Social Science Research Network. <https://papers.ssrn.com/abstract=1616460>

Bason, C. (2010). *Leading Public Sector Innovation: Co-creating for a Better Society*. Policy Press.

Bason, C. (2017). *Leading Public Design. Discovering Human-Centred Governance* (Policy Press. University of Bristol). <https://policy.bristoluniversitypress.co.uk/leading-public-design>

Befani, B. (2016). *Choosing Appropriate Evaluation Methods. A Tool for Assessment & Selection*. Bond. <https://www.bond.org.uk/resources/evaluation-methods-tool>

Bentzen, T. Ø., Sørensen, E., & Torfing, J. (2020). Strengthening public service production, administrative problem solving, and political leadership through co-creation

of innovative public value outcomes? *The Innovation Journal: The Public Sector Innovation Journal*, 25(1).

Berg-Schlosser, D., & Cronqvist, L. (2005). Macro-Quantitative vs. Macro-Qualitative Methods in the Social Sciences—An Example from Empirical Democratic Theory Employing New Software“. *Historical Social Research / Historische Sozialforschung*, 30(4 (114)), 154-175.

Bhatta, G. (2003). *Revisiting the Issue of Risks in Innovation in the Public Sector*. 8, 12.

Bingham, L. B., Nabatchi, T., & O’Leary, R. (2005). The New Governance: Practices and Processes for Stakeholder and Citizen Participation in the Work of Government. *Public Administration Review*, 65(5), 547-558.

Bommert, B. (2010). Collaborative Innovation in the Public Sector. *International Public Management Review*, 11(1), Article 1.

Bourgon, J. (2011). *A New Synthesis of Public Administration: Serving in the 21st Century*. School of Policy Studies, Queen’s University.

Brandsen, T. (2018). Co-Production and Co-Creation: Engaging Citizens in Public Services. *Co-Production and Co-Creation Engaging Citizens in Public Services*. https://www.academia.edu/39272920/Co_Production_and_Co_Creation_Engaging_Citizens_in_Public_Services

Bryson, J., & Crosby, B. (2014). Public Value Governance: Moving Beyond Traditional Public Administration and the New Public Management. *Public Administration Review*, 74(4). <https://doi.org/10.1111/puar.12238>

Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. Palgrave Macmillan UK. <https://doi.org/10.1057/978-1-137-51781-4>

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

Carey, G., & Crammond, B. (2015). What Works in Joined-Up Government? An Evidence Synthesis. *International Journal of Public Administration*, 18, 1-10. <https://doi.org/10.1080/01900692.2014.982292>

Crosby, B., Hart, P., & Torfing, J. (2016). Public value creation through collaborative innovation. *Public Management Review*, 19, 1-15. <https://doi.org/10.1080/14719037.2016.1192165>

Crowley, K., & Head, B. W. (2017). The enduring challenge of ‘wicked problems’: Revisiting Rittel and Webber. *Policy Sciences*, 50(4), 539-547. <https://doi.org/10.1007/s11077-017-9302-4>

Dahler-Larsen, P. (2012). *The Evaluation Society*. Stanford University Press.

Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies*, 47(2), 108-121. <https://doi.org/10.1111/1467-8527.00106>

de Marchi, G., Lucertini, G., & Tsoukias, A. (2016). From evidence-based policy making to policy analytics. *Annals of Operations Research*, 236(1), 15-38. <https://doi.org/10.1007/s10479-014-1578-6>

Doz, Y., & Koskonen, M. (2014). *Governments for the future: Building the strategic and agile state*. Sitra. <https://media.sitra.fi/2017/02/23222725/Selvityksia80.pdf>

Dunleavy, P., & Hood, C. (1994). From old public administration to new public management. *Public Money & Management*, 14(3), 9-16. <https://doi.org/10.1080/09540969409387823>

Emerson, K., Nabatchi, T., & Balogh, S. (2012). An Integrative Framework for Collaborative Governance. *Journal of Public Administration Research and Theory*, 22(1), 1-29. <https://doi.org/10.1093/jopart/mur011>

Fischer, F. (1993). Citizen Participation and the Democratization of Policy Expertise: From Theoretical Inquiry to Practical Cases. *Policy Sciences*, 26(3), 165-187.

Fischer, F., & Forester, J. (1993). *The argumentative turn in policy analysis and planning*. UCL. <http://www.myilibrary.com?id=40672>

Furubo, J.-E., Rist, R., & Sandahl, R. (2002). *International Atlas of Evaluation*. Transaction Publishers.

Geyer, R., & Rihani, S. (2010). *Complexity and Public Policy: A New Approach to 21st Century Politics, Policy And Society*. Routledge. <https://www.routledge.com/Complexity-and-Public-Policy-A-New-Approach-to-21st-Century-Politics-Policy/Geyer-Rihani/p/book/9780415556637>

Goldsmith, S., & Eggers, W. (2004). The new shape of government. En *The New Shape of Public Sector*. Brookings Institution Press.

Hartley, J. (2005). Innovation in Governance and Public Services: Past and Present. *Public Money & Management*, 25(1), 27-34. <https://doi.org/10.1111/j.1467-9302.2005.00447.x>

Head, B. W. (2022). *Wicked Problems in Public Policy: Understanding and Responding to Complex Challenges*. Palgrave Macmillan.

Healey, P. (1997). Systemic Institutional Design for Collaborative Planning. En P. Healey (Ed.), *Collaborative Planning: Shaping Places in Fragmented Societies* (pp. 284-314). Macmillan Education UK. https://doi.org/10.1007/978-1-349-25538-2_9

Henry, G., & Mark, M. (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)

Hood, C. (1995). The "new public management" in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2-3), 93-109. [https://doi.org/10.1016/0361-3682\(93\)E0001-W](https://doi.org/10.1016/0361-3682(93)E0001-W)

Hornby, P., & 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>

Howlett, M. (2014). From the 'old' to the 'new' policy design: Design thinking beyond markets and collaborative governance. *Policy Sciences*, 47(3), 187-207. <https://doi.org/10.1007/s11077-014-9199-0>

- Howlett, M. (2019). *Designing Public Policies: Principles and Instruments* (2.^a ed.). Routledge. <https://doi.org/10.4324/9781315232003>
- Innes, J. E., & Booher, D. E. (1999). Consensus Building and Complex Adaptive Systems. *Journal of the American Planning Association*, 65(4), 412-423. <https://doi.org/10.1080/01944369908976071>
- Jacob, S., Speer, S., & Furubo, J.-E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21(1), 6-31. <https://doi.org/10.1177/1356389014564248>
- Khagram, S., Thomas, C., Lucero, C., & Mathes, S. (2009). Evidence for development effectiveness. *Journal of Development Effectiveness*, 1(3), 247-270.
- King, A., & Crewe, I. (2014). *The Blunders of Our Governments*. Simon and Schuster.
- Mak, S., & Thomas, A. (2022). Steps for Conducting a Scoping Review. *Journal of Graduate Medical Education*, 14(5), 565-567. <https://doi.org/10.4300/JGME-D-22-00621.1>
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs* (1st ed). Jossey-Bass. <http://catdir.loc.gov/catdir/toc/onix06/00010232.html>
- Mead, L. M. (2015). Only connect: Why government often ignores research. *Policy Sciences*, 48(2), 257-272. <https://doi.org/10.1007/s11077-015-9216-y>
- Mello, P. (2021). *Qualitative Comparative Analysis: An Introduction to Research Design and Application*. Georgetown University Press. <https://patrickmello.com/qca-research-design-application/>
- Meyer, W., Stockmann, R., & Taube, L. (2020). The Institutionalisation of Evaluation Theoretical Background, Analytical Concept and Methods. En R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 3-34). https://doi.org/10.1007/978-3-030-32284-7_1
- Mulgan, G. (2014). *Innovation in the public sector. How can public organisations better create, improve and adapt?* NESTA. https://media.nesta.org.uk/documents/innovation_in_the_public_sector-how_can_public_organisations_better_create_improve_and_adapt_0.pdf
- Nambisan, S. (2008). *Transforming Government Through Collaborative Innovation* (p. 44). IBM Center for the Business of Government.
- Newman, J., & Head, B. (2015). Beyond the two communities: A reply to Mead's «why government often ignores research». *Policy Sciences*, 48(3), 383-393.
- Nie, M. (2003). Drivers of natural resource-based political conflict. *Policy Sciences*, 36(3), 307-341. <https://doi.org/10.1023/B:OLIC.0000017484.35981.b6>
- OECD. (2001). *Citizens as Partners: Information, Consultation and Public Participation in Policy-Making*. OECD Publishing. https://www.oecd.org/en/publications/citizens-as-partners_9789264195561-en.html
- OECD. (2020). *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. Organisation for Economic Co-operation and Development.

https://www.oecd-ilibrary.org/governance/improving-governance-with-policy-evaluation_89b1577d-en

Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation*, 22(2), 168-189. <https://doi.org/10.1177/1356389016638752>

Osborne, S. P. O. (2006). The New Public Governance? *Public Management Review*, 8(3), 377-387. <https://doi.org/10.1080/14719030600853022>

Paquet, G. (2009). *Scheming virtuously: The road to collaborative governance*. Invenire Books.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use* (pp. xxi, 375). Guilford Press.

Pattyn, V. (2014). *Policy evaluation (in)activity unravelled. A configurational analysis of the incidence, number, locus and quality of policy evaluations in the Flemish public sector (2007-2014)* [Page, KU Leuven]. <https://soc.kuleuven.be/io/english/research/project/doctoral-policy-evaluation-unravelled>

Pattyn, V., & Bouterse, M. (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>

Pattyn, V., Voorst, S., Mastenbroek, E., & Dunlop, C. (2017). Policy Evaluation in Europe. En *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 577-593). https://doi.org/10.1057/978-1-137-55269-3_30

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. SAGE Publications.

Ragin, C. C. (1989). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.

Rihoux, B., & Ragin, C. 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>

Rihoux, B., Rezsöházy, I., & Bol, D. (2011). Qualitative Comparative Analysis (QCA) in Public Policy Analysis: An Extensive Review. *German Policy Studies*, 7.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155-169. <https://doi.org/10.1007/BF01405730>

Ryan, M., & Smith, G. (2013). Towards a Comparative Analysis of Democratic Innovations. Lessons from a small-N fsQCA of Participatory Budgeting. *Revista Internacional de Sociología*, 70, 89-120. <https://doi.org/10.3989/ris.2012.01.28>

Sørensen, E., & Torfing, J. (2011). Enhancing Collaborative Innovation in the Public Sector. *Administration & Society*, 43(8), 842-868. <https://doi.org/10.1177/0095399711418768>

Sorrentino, M., Sicilia, M., & Howlett, M. (2018). Understanding co-production as a new public governance tool. *Policy and Society*, 37(3), 277-293. <https://doi.org/10.1080/14494035.2018.1521676>

Stockmann, R., & Meyer, W. (2022a). The Institutionalisation of Evaluation in Europe and the Americas: A Comparison. En R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in the Americas* (pp. 509-525). Springer International Publishing. https://doi.org/10.1007/978-3-030-81139-6_17

Stockmann, R., & Meyer, W. (2022b). The Institutionalisation of Evaluation in the Americas: A Synthesis. En R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in the Americas* (pp. 451-507). Springer International Publishing. https://doi.org/10.1007/978-3-030-81139-6_16

Stockmann, R., Meyer, W., & Szentmarjay, L. (Eds.). (2022). *The Institutionalisation of Evaluation in the Americas*. Palgrave Macmillan Cham. <http://link.springer.com/book/10.1007/978-3-030-81139-6>

Stockmann, R., Meyer, W., & Taube, L. (2020a). The Institutionalisation of Evaluation in Europe: A Synthesis. En R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 483-522). Springer International Publishing. https://doi.org/10.1007/978-3-030-32284-7_19

Stockmann, R., Meyer, W., & Taube, L. (Eds.). (2020b). *The Institutionalization of Evaluation in Europe* (Vol. 34). Palgrave Macmillan Cham.

Stockmann, R., Meyer, W., & Zierke, N. (Eds.). (2023). *The Institutionalisation of Evaluation in Asia-Pacific*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-36918-6>

Stoker, G. (2006). Public Value Management: A New Narrative for Networked Governance? *The American Review of Public Administration*, 36(1), 41-57. <https://doi.org/10.1177/0275074005282583>

Straus, D. (2002). *How to Make Collaboration Work: Powerful Ways to Build Consensus, Solve Problems, and Make Decisions*. Berrett-Koehler.

Torring, J. (2016). *Collaborative Innovation in the Public Sector*. Georgetown University Press. <https://www.jstor.org/stable/j.ctt1h64pfm>

Torring, J., Andersen, L. B., Greve, C., & Klausen, K. K. (2020). New Public Governance. En *Public Governance Paradigms* (pp. 125-150). Edward Elgar Publishing. http://www.elgaronline.com/display/9781788971218/13_chapter8.xhtml

Torring, J., Ferlie, E., Jukić, T., & Ongaro, E. (2021). A theoretical framework for studying the co-creation of innovative solutions and public value. *Policy & Politics*, 49(2), 189-209. <https://doi.org/10.1332/030557321X16108172803520>

Torring, J., Røiseland, A., & Sørensen, E. (2016, agosto 26). *Transforming the Public Sector into an Arena for Co-creation: Barriers, Drivers, Benefits and Ways Forward*.

Uusikylä, P., Ahonen, P., Hietakangas, N., & Nurmi, J. (2020). *Finland* (pp. 63-86). https://doi.org/10.1007/978-3-030-32284-7_3

Varone, F., & Jacob, S. (2004). Institutionnalisation de l'évaluation et nouvelle gestion publique: Un état des lieux comparatif. *Revue internationale de politique comparée*, 11(2), 271-292. <https://doi.org/10.3917/ripc.112.0271>

Varone, F., Rihoux, B., & Marx, A. (2006). *A new method for policy evaluation? Longstanding challenges and the possibilities of qualitative comparative analysis (QCA)* (pp. 213-236). https://doi.org/10.1007/0-387-28829-5_10

Wagenaar, H. (2007). Governance, Complexity, and Democratic Participation: How Citizens and Public Officials Harness the Complexities of Neighborhood Decline. *The American Review of Public Administration*, 37(1), 17-50. <https://doi.org/10.1177/0275074006296208>

Webber, M. M. (1983). The Myth of Rationality: Development Planning Reconsidered. *Environment and Planning B*, 10(1), 89-99.

Weber, E. P., & Khademian, A. M. (2008). Wicked Problems, Knowledge Challenges, and Collaborative Capacity Builders in Network Settings. *Public Administration Review*, 68(2), 334-349. <https://doi.org/10.1111/j.1540-6210.2007.00866.x>

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

2. Mechanisms for institutionalising evaluation: a scoping review

Author: Carla Cordoncillo Acosta

Publication state: Under review after a resubmission for major changes in *Evaluation and Program Planning*

Abstract

In recent years, there has been growing attention garnered by the institutionalisation of evaluation, with academia as well as international organisations, governments and practitioners engaging more with the topic. Recent publications, in particular the books edited by Stockmann and Meyer (2020, 2022, 2023) that bring together various experiences occurring in different European, American and Asian countries, have contributed significantly to the conceptual and theoretical development of the field. However, the predominant analytical frameworks used to assess institutionalisation are primarily designed to support international comparisons and to quantify the degree of institutionalisation across countries. As such, they tend to emphasise measurable indicators or enabling conditions, often overlooking the specific mechanisms and institutional arrangements that underpin the development and sustainability of evaluative practices. This article addresses that gap by conducting a scoping review of 29 case studies from 12 countries with different evaluation traditions. Rather than focusing on levels of institutionalisation, the analysis identifies and categorises the institutional arrangements and mechanisms most frequently used to embed evaluation within public administration. In doing so, it offers a structured overview intended to support public sector managers—particularly in contexts with limited evaluation traditions—in reflecting on and designing appropriate strategies to strengthen evaluation systems.

2.1 Introduction

Although it might be argued that evaluation is as old as human activity —because the act of weighing up alternatives when faced with a problem and taking decisions based on that evaluation is surely as old as reason (Shadish & Luellen, 2005)— evaluation as an academic discipline is still relatively young. Although its focus was originally on methodological aspects and the use of its results (Alkin et al., 1979; Alkin & Daillak, 1979; M. Patton, 1988b, 1988a; M. Q. Patton et al., 1977; C. Weiss, 1979; C. H. Weiss, 1988), “the characteristics of individual evaluations cannot explain why some institutions, regions and countries tend to systematically produce higher quality evaluations, which are better integrated into the decision-making process” (Martinaitis et al., 2019:1). As a result, in recent decades scholars turned their interest towards the “evaluation capacity” (EC) and the “evaluation capacity development” (ECB) of organisations (Bohni Nielsen et al., 2011; Bourgeois & Cousins, 2013; Compton et al., 2002; Labin et al., 2012; Milstein & Cotton, 2000; Preskill & Boyle, 2008; Segalovičienė, 2012; Stockdill et al., 2002). They have subsequently considered evaluation systems (Goldman et al., 2023; Leeuw &

Furubo, 2008; Martinaitis et al., 2019; Raimondo, 2018) and more recently the culture of evaluation, and above all, **its institutionalisation**.

In recent years, case studies and comparative analysis on the institutionalisation of evaluation have proliferated, largely due to the books edited by Stockmann and Meyer (2020, 2022, 2023). The principal analytical frameworks employed to examine the institutionalisation of evaluation (Furubo et al., 2002; Jacob et al., 2015a; Meyer et al., 2020; OECD, 2020; Varone & Jacob, 2004) are designed not only to facilitate comparisons, but, in particular, to quantify the degree of institutionalisation across countries. Consequently, these frameworks tend to prioritise indicators (e.g., the frequency of evaluations) or conditions (e.g., the extent to which evaluation occurs across multiple policy domains) that enable the operationalisation of institutionalisation, rather than focussing on the underlying mechanisms. **However, to develop a more comprehensive understanding of the strategies that countries can adopt to institutionalise policy evaluation, it is essential to move beyond indicators and instead identify the institutionalisation mechanisms and arrangements most commonly utilised.** While certain nuances between these concepts can be acknowledged, in this article the terms will be used interchangeably to refer to the structures, strategies, processes, and tools that are presumed to influence the regularity of evaluations and their embedding within public institutions.

To undertake the task of identification and categorisation, a scoping review was conducted focusing on mechanisms within political systems that institutionalise evaluation. The review analysed literature on the institutionalisation of evaluation in 12 countries with different evaluation traditions. Allowing for the systematic identification and classification of the specific institutional arrangements employed in each case. As Jacob (2005b, p. 51) notes, categorisations facilitate the “ordering features into mutually exclusive and collectively exhaustive categories”. Given that the institutionalisation of evaluation serves the public interest and falls within the remit of public administrations, the review was designed with a practical orientation. It identifies key factors that public managers should consider when selecting institutional arrangements, aligned with the principal decisions required to establish a national evaluation system: defining the purpose of evaluation, determining what will be evaluated, identifying who will conduct evaluations, specifying under what conditions evaluation will occur, and outlining how results will inform decision-making processes. These factors are framed as guiding questions intended to support public managers in choosing the most appropriate mechanisms. By offering a comprehensive, systematically organised list of institutionalisation mechanisms, and by providing concrete elements for consideration in their selection, the study aims both to advance the literature on the institutionalisation of evaluation and to inform decision-making processes within public administrations.

2.2 Analytical framework

Evaluation culture and institutionalisation of evaluation are two concepts that have often been **used interchangeably**, even in reference works (Furubo et al., 2002; Jacob et al., 2015b), but they are not synonyms (Jacob, 2023a). In fact, Jacob (2023) identifies this

terminological confusion as one of the limitations of the literature on the institutionalisation of evaluation, and highlights the need to clarify the difference between the two terms in order to systematise future research in this area

Evaluation culture is a concept that refers to an organisation's commitment to the role of evaluation in decision-making processes (Owen & McDonald, 1999), and therefore emphasises what Leeuw and Furubo (2008) call the epistemological dimension of evaluation systems. As Patton noted at the end of the last century, this concept is linked to the notion of "organisational culture", where organisation can refer to anything from a specific public agency to the entire state administration. According to Kim (2002:3, quoting Cameron and Quinn, 1999), organisational culture "is reflected by what is valued, the dominant leadership styles, symbols, the procedures, routines, and the definition of success that make an organization unique". In this sense, Dahler-Larsen and Boodhoo (2019: 280) consider that "evaluation culture describes evaluative mindsets and how people make evaluation fit into local beliefs and traditions".

These definitions align with what Schein (2010) describes as the second and third "levels" through which an organisation's culture can be observed: the values it adopts (its priorities, philosophy, and strategies) and the premises or assumptions underlying those values, i.e. the unconscious beliefs, perceptions and thoughts at play. The first level, however, refers to what the author calls "artifacts" — defined as visible organisational processes, structures and behaviours— a concept and definition that is very close to that of "institutional devices" for evaluation, used by Jacob (2005:838) to refer to the set of "organizations and rules that contribute to the development and sustainability of the evaluative practice". In fact, according to Pattyn (2015), **institutionalising evaluation** precisely means developing institutional arrangements with the intention of achieving regular evaluative practices. Therefore, **institutionalisation is the visible level (the first level) of evaluation culture**.

The fact that institutionalisation represents the most visible aspect of evaluation culture makes it also the most concrete and easily observable. This explains why the bulk of recent literature on the subject, especially in terms of comparative studies —both quantitative and qualitative — focus on this concept. In fact, Dahler-Larsen and Boodhoo (2019:281) attribute a second meaning to the concept of evaluation culture that refers precisely to "the comparative degree of institutionalisation of evaluation." However, while this might add to the confusion between the two concepts, their explanation links the existence of institutional arrangements related to evaluation to "evaluative maturity". Thus, although they do not do so explicitly, they imply that an **organisation or administration that has *matured* in terms of evaluation** is one that has a high degree of institutionalisation of evaluation (which can be operationalised through indices that reflect the availability of institutional arrangements). They do however differentiate this evaluative maturity from a culture of evaluation, since those indices establish a linear and unique conception of success, and do not consider the context (normative, administrative, or organisational), which is essential when we talk about a culture (Dahler-Larsen and Boodhoo, 2019).

In this sense, it is interesting that, as Jacob (2023) says, the indices that exist to measure how institutionalised evaluation is in a country —i.e. in the state administration— **all**

bring together very similar elements, despite having been developed by a number of different researchers. As Table 1 shows, the proposals list mechanisms related both to the political system —i.e., the institutional structures and decision-making processes of the executive and legislative branches (Meyer et al., 2020) and to the system of professionalisation —i.e., the full range of organisations that offer evaluation services and training. All this reflects what has traditionally been addressed as demand and supply of evaluation (Toulemonde, 1999). However, Meyer et al. (2020) incorporate institutionalisation within what they call the “social system”, referring to the dissemination and acceptance of evaluation among civil society.

In this way, **quantitative comparative studies allow countries to be ranked according to their evaluative maturity**, since "the assumption underlying these studies is that the more institutional mechanisms and arrangements are distributed within the state apparatus, the more embedded evaluations become and the more they can fulfil their purposes in terms of accountability and improved public management" (Jacob, 2023:188). Most qualitative studies, on the other hand, are case studies that conduct longitudinal analyses to delve more deeply into the particular situation of each country or region.

Table 1: Factors involved in institutionalising evaluation

Furubo et al. (2002), Jacob et al. (2015)	Varone and Jacob (2004)	Meyer et al. (2020)
Organisations and evaluation structure		
Evaluation takes place in many policy domains.		Sectoral spread, scope and frequency of evaluation (all policy sectors or one field in particular).
Institutional arrangements within governments to conduct evaluations and to disseminate their results to decision makers.	Institutional arrangements within governments to conduct evaluations.	Existence of specific evaluation units in government institutions (such as ministries) and/or existence of independent stand-alone evaluation bodies.
Institutional arrangements are present in parliament to conduct evaluations and disseminate their results to decision makers.	Institutional arrangements present in parliament to conduct evaluations.	Parliament has an evaluation unit at its disposal and commissions evaluations.
Evaluation within the supreme audit institution.	Evaluation within the supreme audit institution..	National audit offices carry out not only performance audits (limited to evaluating achievement and/or efficiency), but also evaluations with a broader focus.
Rules, forum, practices		
		National or sectoral laws, decrees or regulations about evaluation.
Supply of domestic evaluators from various disciplines.		
Professional Organisation (VOPE*)	Professional Organisation (VOPE*)	Professional Organisation (VOPE*)

National discourse concerning evaluation.		
Plurality of institutions or evaluators performing evaluations within each policy domain.		
	Scientific journal	Specialised scientific journals.
		Education - academic and non-academic training programmes.
	Quality standards	Obligatory standards and/or certifications.
Use of evaluation		
Consideration of the effects and implementation of public policies through evaluation.		Parliament regularly takes note of and discusses evaluation results. Institutionalised use of evaluation in the social system (by civil society).

Source: Jacob (2023). *VOPE: voluntary organisation for professional evaluation.

The combination of strategies and mechanisms deployed to institutionalise evaluation in each context constitutes its **evaluation system**. According to Leeuw and Furubo (2008), however, for a set of evaluation mechanisms and activities to be considered an evaluation system, four criteria must be met: (1) there must be a shared epistemology that allows us to identify what can be considered evaluation; (2) the evaluation practice must be driven by organisations, and not by individuals; (3) the mechanisms and activities developed must have a largely permanent nature; (4) there must be some kind of link in place between the information derived from the evaluations carried out and subsequent decision-making processes (i.e. an explicit will to use the results of the evaluations). Goldman et al. (2023:317) sum this up, defining evaluation systems as a "set of rules and processes that are carried out in a systematic way to coordinate, develop and use evaluations".

In this way, **evaluation systems evolve as evaluation becomes more institutionalised**. According to Jacob (2005a), the institutionalisation of evaluation **is an evolutionary process that works in an adaptive and aggregative way**. First, a mechanism is designed and implemented; second, an assessment is made of how it operates and how well it fits with the other mechanisms in the system; third, it is modified according to the appraisals given by the actors involved. Fourth, if it is seen as viable, when it has eventually been consolidated, the process begins all over again with a new mechanism, which is designed and eventually implemented. In this way, administrations build institutional environments that work to consolidate evaluation practices and help align and systematise evaluation efforts and activities that until that moment have been sporadic or dispersed (Gaarder & Briceño, 2010).

The numerous comparative analyses on the institutionalisation of evaluation (Jacob et al., 2015b; Lázaro, 2015; OECD, 2020; Stockmann & Meyer, 2022a, 2022b) underline that there is no single path to achieving a mature evaluation system. Instead, it is clear that each system—as well as its evolution—is the result of countless factors: the legal and administrative culture (Jacob, 2023a), the motivation to evaluate (Goldman et al.,

2023), the political context (Chelimsky, 2009), the role that scientific knowledge has traditionally garnered in public policy (Jacob, 2023a), the influence of international organisations, such as the European Union or the United Nations (Jacob, 2023a), to name but a few. For this reason, in a comparative perspective, it is important to move beyond the quantitative assessment of evaluation maturity and to advance the understanding of the range of institutionalisation mechanisms and the factors relevant to their selection. The analysis presented below seeks to support this endeavour by contributing to the construction of institutional environments that are favourable to evaluation, particularly in contexts which have very little tradition of evaluation or where evaluation practices are not yet well established.

2.3 Methodology

A scoping review methodology has been used to systematically examine the existing literature on the institutionalisation of evaluation and to answer the following question: **what are the institutional mechanisms and arrangements currently in place to ensure that evaluation is routinely employed?** The method followed the steps used for scoping reviews identified by both the Joanna Briggs Institute (JBI) and Arksey and O'Malley (2005): (1) the research question was developed; (2) the relevant studies were identified; (3) a selection was made of the studies; (4) data extraction was performed; (5) a descriptive analysis of the studies was carried out and (6) there was a discussion of the results.

2.3.1 Identification of relevant studies

After a series of exploratory searches using keywords, the final search for articles was carried out by using two databases (Scopus and Web of Science) and by employing the following terms: "institutionalization of evaluation" OR "institutionalisation of evaluation" OR "culture of evaluation" OR "evaluation culture" OR "institutionalization of policy evaluation" OR "institutionalisation of policy evaluation" OR "institutionalising evaluation" OR "institutionalizing evaluation" OR "institucionalización de la evaluación" OR "cultura de la evaluación" OR "institutionnalisation de l'évaluation" OR "culture de l'évaluation". The following eligibility criteria were applied: (1) academic articles, doctoral theses or book chapters, (2) published in English, Spanish or French, (3) with no time limitations (since this is a relatively new field of knowledge), (4) specifically addressing the institutionalisation of evaluation in the political and/or professionalising system (thereby covering supply and demand, but leaving out the social system), (5) in the following 12 countries: Sweden, Canada, United Kingdom, Australia, Netherlands, Finland, Spain, Chile, Switzerland, Mexico, Romania and Korea.

Following Vedung (2010), three countries were selected to reflect each of the four waves of evaluation diffusion. Germany, Canada and the United Kingdom were placed in the first wave, which ran from the late 1950s to the late 1960s and was marked by the consolidation of welfare states and a radical rationalism that advocated applying scientific principles to public policies (Vedung, 2010). Australia, the Netherlands and Finland were selected for the second wave, characterised by the erosion of the positivist

paradigm in favour of constructivist approaches, in an international context of political and economic instability (the 1970s). Spain, Chile and Switzerland were chosen for a third (neoliberal) wave, clearly marked by the rise of New Public Management and the public policy evaluation requirements that the European Union demanded of its members. And finally, Mexico, Romania and Korea were selected as countries belonging to the fourth wave, the wave that we have found ourselves in since the turn of the century, although some scholars defend that a fifth wave also exists (Picciotto, 2015) which is characterised by the return of experimentation, hand in hand with evidence-based policies.

All the countries analysed have national evaluation systems that encapsulate their institutionalisation initiatives, which means that no analysis is carried out here regarding the existing mechanisms in contexts where institutionalisation initiatives are still spontaneous or isolated.. The countries selected for each wave have been chosen considering the existing academic research and seeking to guarantee the maximum territorial heterogeneity. At the same time, including countries with different evaluation traditions guarantees a greater plurality of institutionalisation strategies.

2.3.2 Selection and coding of studies

The studies were selected in three phases, a procedure recommended by the Joanna Briggs Institute (Vedung, 2010). First, a search was carried out by applying the eligibility criteria number 1, 2, 3 and 5, resulting in **159 studies** being identified. Microsoft Excel was then used to organise the studies and eliminate duplicates, which resulted in a **total of 144 studies**. Finally, criterion 4 was applied (and criterion 5 was revised) based on reading the abstracts, leaving us with **31 articles** to be analysed. Five articles were also identified that included analyses that were either continental or related to several countries, and nine theoretical articles were found on the institutionalisation of evaluation in the political system, but without reference to a specific territory. These 14 additional articles have been taken into account for the discussion but have not been included in the analysis itself.

After a thorough reading of the 31 articles, two were discarded for not meeting criterion 4. The **remaining 29** were considered for the scoping review and had the information collected in Table 2 extracted from them. Figure 1 summarises the process of searching for and selecting the studies and Appendix 1 details the items from Table 2 for each of the articles included in the scoping review.

Table 2. Information extracted from the studies analysed

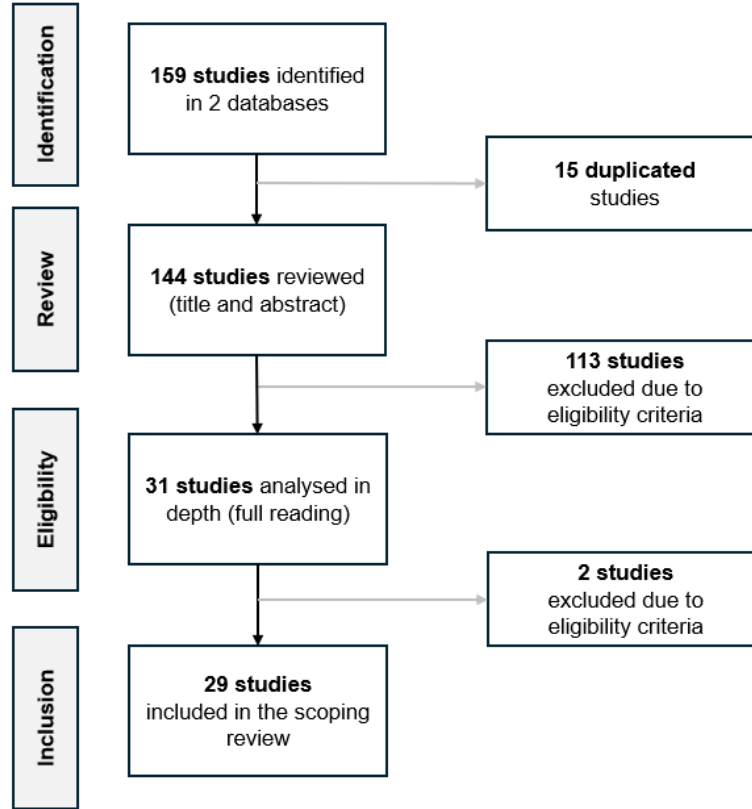
General information about the study

- Author
- Year of publication
- Type of document (academic article, doctoral thesis or book chapter)

Geographical information (in terms of institutionalising evaluation)

- Country referred to in the study
- Wave (of evaluation becoming diffused)
- Administrative level of the administration analysed: state, regional, local

Figure 1. PRISMA diagram of the search and selection process of the studies



In order to extract and code the various mechanisms, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses – Extension for Scoping Reviews (PRISMA-ScRP) system was used. Both its checklist (Tricco et al., 2018) as well as the qualitative information extraction tool proposed by the JBI (Aromataris et al., 2024) were employed. The information was analysed using qualitative thematic analysis, resulting in a categorisation of the mechanisms employed, as well as a conceptual and practical discussion presented in the section below.

2.4 Results

2.4.1 Characteristics of the studies

Of the 29 studies analysed, six refer to countries belonging to the American continent, 18 to European countries (although one of them analyses two countries), four to Australia (Oceania) and only one study analyses an Asian country, Korea. Table 3 shows the frequency of studies found for each of the 12 countries considered in the search.

Table 3: Number of studies considered in the analysis by country, continent and wave of when evaluation began to be diffused

Continent/Country		Number of studies
America		6
Canada	(1st wave)	2
Chile	(2nd wave)	1
Mexico	(4th wave)	3
Asia		1
Korea	(4th wave)	1
Europe		18
Germany	(1st wave)	1
Spain	(3rd wave)	5
Finland	(2nd wave)	2
Netherlands	(2nd wave)	2
United Kingdom	(1st wave)	1
Romania	(4 wave)	2
Switzerland	(3rd wave)	6
Oceania		4
Australia	(2nd wave)	4
Grand total		30¹

It can be seen that the majority of studies are concentrated in countries belonging to the second (8 studies) and third (12 studies) waves of evaluation diffusion. It is also important to highlight that most of the studies (23 out of 29) concentrate on the **state level**, as this is typically the administrative level where the institutionalisation of evaluation is most advanced and therefore, the level on which academic research has focused. Finally, six of the articles analysed were published between 2005 and 2009, nine between 2010 and 2019 and 15 between 2020 and 2023, and almost half are book chapters. The volume of publications in recent years is explained by the compilations edited by Stockmann et al. (2020, 2022, 2023), highlighting their important contribution to research on the institutionalisation of evaluation.

2.4.2 Identification and categorisation of institutionalisation strategies in political system

The indices operationalising the institutionalisation of evaluation presented in Table 1 distinguish between: (1) evaluation legislation, (2) evaluation structures, (3) evaluation practices and (4) uses of evaluation. The OECD (2020), however, in the report resulting of its Survey on Institutionalisation, Quality and Use of Policy Evaluation, considers an additional dimension: (5) the evaluation policy framework. These categories are then used to identify and organise the institutionalisation mechanisms inferred from the scoping review.

¹The total is 30 studies (instead of 29) because one study analyses two cases: the Netherlands and Switzerland.

Evaluation legislation

"The embedding of evaluation within legislation often represents a turning point in the process of institutionalization" (Jacob, 2023:197). This embedding, however, can occur in different levels of legislation, something that could be considered an indicator of the relevance given to evaluation, although of course this is determined by the legal framework and legislative tradition of each country. Only Switzerland and Mexico dedicate an article in their Constitutions (the highest-ranking legal norm) to the evaluation of public policies. In this sense, it is worth noting that Horber-Papazian and Baud-Lavigne (2019) consider this incorporation of the obligation to evaluate in the national Constitution a determining factor in Switzerland's evaluation maturity (Furubo et al., 2002; Jacob et al., 2015b).

Other countries (among those analysed, only Spain) have opted for passing a specific law on the institutionalisation of evaluation. The most common practise, however, is to address the issue with one or more sectoral laws (this had been the case in Spain until 2022), among which the most notable are laws relating to the national budget or public finances (e.g. Australia, Chile, Finland and the Netherlands). Finally, there are countries that regulate evaluation with secondary legislation, such as decrees and resolutions (e.g. Germany or the UK).

Beyond the range of legislation that governs evaluation, there are also other types of regulatory mechanisms for institutionalising evaluation: **requirements**. In fact, in almost all the countries analysed, current legislation requires ex-post evaluation of at least the most strategic policies. In some countries, ex-ante evaluations of new laws or policies are also required, and in Germany, the Netherlands and Switzerland, it is common to include evaluation clauses in legislative proposals.

These types of regulatory mechanisms —evaluation requirements— imply an obligatory nature. According to McDonald et al. (2003), a key aspect when managing the demand for evaluation is to define whether it should be mandatory. According to the authors, "making evaluation mandatory could promote a culture of token compliance, but voluntary adoption is much slower to take effect" (McDonald et al., 2003:11). In line with the first argument, Cardozo (2009:191) mentions that in Mexico, their mandatory nature has led to "evaluations being seen as just one more administrative requirement in relation to budget allocation; an activity that must be fulfilled but is not considered a fundamental working tool to improve design and management".

The way in which these different mechanisms (summarised in Table 3) are combined defines the regulatory framework of evaluation in each context, which in turn directly determines the evaluation practices, especially on the demand side.

Table 4. Mechanisms for institutionalising evaluation: legislation and evaluation requirements

<i>Institutionalisation mechanisms</i>	<i>Number of studies</i>	<i>Number of countries</i>	<i>Countries²</i>
Legislation			
<i>Legislation related to evaluation</i>			
→ Specific article in the country's constitution.	6	2	MX, SW
→ Specific law on evaluation.	2	2	KO, SP
→ Regulation in sectoral laws (especially in public finance or budgetary laws).	9	8	AU, CHI, FIN, KO, MX, NE, SP, SW
→ Regulation in secondary legislation.	4	3	AU, FIN, UK
<i>Evaluation requirements</i>			
→ Ex-ante reports on the analysis of the impact of new regulations or policies ³ .	6	6	CHI, FIN, GER, NE, SW, UK
→ Results-based accountability / Decision-making based on effectiveness and efficiency criteria.	4	4	FIN, GER, NE, SW
→ Evaluation of strategic policies.	18	10	AU, CA, CHI, GER, KO, MX, NE, RU, SP, SW
→ Evaluation clauses in legislative proposals.	7	3	GER, NE, SW

Evaluation structures

According to Jacob (2023:197), in several countries “the development of evaluation is the result of work undertaken by agents of change who occupy a central position in the functioning of the state”. This work, however, can be carried out from very different organisational structures, as evidenced by the results of the scoping review. The first thing to bear in mind is that promoting evaluation does not necessarily mean also carrying it out. A unit may be assigned the task of evaluating but may limit itself to commissioning other agents to carry out the evaluations themselves, while another may choose to conduct them directly. And both these situations stimulate the demand for evaluation. Therefore, when deciding who takes on the task of promoting evaluation, it is important to consider: (1) who can commission evaluations; (2) who carries out evaluations (and what position they have in the institutional organigram, which may determine how independent their results are); (3) who ensures their quality (and if there is a specific unit with this function) and (4) who promotes the institutionalisation of evaluation.

As regards where the demand for evaluation originates, the main agent is in all cases the country's government. The differential element is the **possibility for parliamentarians to commission evaluations** (this is only the case in Korea, Finland,

² For all subsequent tables, the following country abbreviations are used: AU (Australia), CA (Canada), CHI (Chile), FIN (Finland), GER (Germany), KO (Korea), MEX (Mexico), NE (Netherlands), RU (Rumania), SP (Spain), SW (Switzerland), and UK (United Kingdom).

³ Reports that analyse economic and social impacts are considered, so those that focus on the regulatory impact are excluded.

the Netherlands and Switzerland). In any case, a national government —along with the full range of public bodies under its authority— can choose to conduct evaluations internally or to outsource them to specialised entities, either public or private. In some countries one or more public institutions —either generalist or sector-specific— are formally assigned the function of carrying out evaluations. These institutions can be affiliated with the executive branch (as is the case in Australia and Chile), the legislative branch (as in Korea and Switzerland), or may operate independently —a far less common arrangement that, among the countries analysed, is found only in Mexico). However, in the majority of countries, the responsibility for evaluation is decentralised and distributed among various ministries or departments. As a result, most ministries or departments maintain dedicated evaluation units of varying size, which in some cases, consist of only a few individuals tasked with carrying out evaluations (this is the case in 10 out of the 12 countries examined).

The extent to which this is feasible depends largely on the degree to which **the role of the evaluator is institutionally embedded and formally recognised** within each public administration. Developing a formal professional recognition is undoubtedly an institutionalisation strategy, which connects with a more general debate on the professionalisation of evaluation. In fact, according to Jacob and Boisvert (2010:363), “when evaluation is well established in the process of public management, that is, when it achieves a certain maturity, questions on the professionalization of evaluation begin to surge”. The formal recognition of evaluators means creating a specific professional category within the public administration, leading to advantages and disadvantages. Although it can be a way of strengthening evaluation, it can also restrict the diversity in term of evaluators' profiles. In this way, its final effect on evaluation practices is ambiguous. According to the results of the scoping review, this formal professional recognition of evaluators within the public administration only occurs in Canada and Switzerland, although in Australia a competency framework for evaluators has been defined.

In fact, most of the studies analysed put significantly **more emphasis on outsourcing evaluations** than on its formal professional recognition. In practically all countries, it is common practice to outsource public policy evaluations, either by entrusting them to academia (universities and public research institutes, as usually happens in the Netherlands, Korea, Finland or the United Kingdom) or to specialised consultancy teams (as it is frequent in Chile or Romania). What does vary substantially is the weight, influence and type of relationship these evaluating bodies have with the public administration.

In some countries (Australia, Canada, Finland, the Netherlands and in part, Korea) there is a structure in place that assumes a third function associated with evaluation: a review (generally based on meta-evaluations) of the quality and/or independence of the evaluations carried out. In some cases, this is a task that is taken on by public research institutes or independent committees formed by researchers, and in other cases it is performed by specific public bodies. It is important not to confuse this function of reviewing or guaranteeing the quality of evaluations with the auditing carried out by each country's Court of Audit or national Audit Office, which in some cases (Australia, Canada, Finland, the Netherlands and Switzerland) also conduct evaluations.

Finally, it is important to distinguish the aforementioned functions —commissioning evaluations, conducting evaluations, and monitoring their quality— from a fourth, distinct function: **promoting the institutionalisation of evaluation. This responsibility is often undertaken by units within the executive branch** —typically under the authority of the ministries of finance or presidential affairs— as is the case in Australia, Canada, Chile, Finland, the Netherlands, and Spain. This arrangement is logical, as the task primarily involves fostering the systematic integration of evaluation practices throughout the public administration. However, in other countries, this function is not a clearly assigned one (as is the case in Korea, where there is no structure that formally assumes the task of institutionalising evaluation).

Table 5. Mechanisms for institutionalising evaluation: structures

<i>Institutionalisation mechanisms</i> Structures	<i>Number of studies</i>	<i>Number of countries</i>	<i>Countries</i>
<i>Function: promoting the institutionalisation and the culture of evaluation</i>			
→ Specific unit(s) dependent on the executive branch.	8	7	AU, CA, CHI, FIN, MX, NE, SP
→ Specific unit dependent on the legislative branch.	4	1	SW
→ Independent unit/agency.	2	1	MX
<i>Function: performing evaluations</i>			
→ Specific unit(s) dependent on the executive branch.	2	2	AU, CHI
→ Specific unit dependent on the legislative branch	5	2	KO, SW
→ Independent unit/agency.	3	1	MX
→ Court of Audit or National Audit Office.	5	5	CA, FIN, NE, SW, UK
→ Specific people or unit(s) within each department.	17	10	AU, CA, FIN, KO, MX, NE, RU, SP, SW
→ Research institutes.	8	6	AU, FIN, KO, NE, SW, UK
<i>Function: reviewing evaluation quality and/or independence</i>			
→ Specific unit(s) dependent on the executive branch.	2	2	AU, CA,
→ Independent bodies, research centres/groups of researchers.	3	3	FIN, KO, NE
→ Meta-evaluations.	3	2	FIN, NE

Evaluation policy framework

Evaluation governance has to do with the rules that regulate evaluation and the structures from which it is promoted, but also with what the OECD (2020) calls the “policy framework” for evaluation. While the Survey on Institutionalisation, Quality and Use of Policy Evaluation (OECD, 2020) only considers the existence of national (Canada, Korea, the Netherlands and the United Kingdom) or sectoral (Australia, Finland, the Netherlands) evaluation strategies, this scoping review allows us to identify other increasingly widespread mechanisms, among which evaluation plans stand out (governmental, sectoral and/or regulatory). “Adopting a multi-year evaluation agenda

helps to prioritize what to evaluate when resources are scarce and when capabilities are burgeoning” (Jacob, 2023:197, quoting Goldman et al., 2018).

On another note, some countries calculate figures for public expenditure on evaluation (Canada), i.e. the total amount of resources (in absolute numbers or as a percentage of the national budget) dedicated to the evaluation of public policies; others calculate the proportion of the public budget that has been subject to evaluation (Canada, Chile, Mexico). This implies first identifying which public policies have been evaluated each year, and then calculating what percentage of the national budget is dedicated to those policies.

Table 6. Mechanisms for institutionalising evaluation: policy framework

<i>Institutionalisation mechanisms</i>	<i>Number of studies</i>	<i>Number of countries</i>	<i>Countries</i>
Policy framework			
<i>Evaluation strategies</i>			
→ National evaluation strategy.	6	5	CA, KO, MX, NE, UK
→ Sectoral evaluation strategy.	3	3	AU, FIN, NE
<i>Evaluation planning (at a systemic level)</i>			
→ Evaluation plans.	10	5	AU, FIN, KO, MX, SP
→ Calculation of the proportion of the national public budget allocated to evaluation.	2	1	CA
→ Calculation of the proportion of the national public budget that has been evaluated.	4	4	CA, CHI, MX, (NE)*

*From the text we can infer that it was a punctual exercise.

Evaluation practices

While all the above strategies should result in an increase in the demand for evaluation and therefore also an increase in evaluation practice, it is also possible to develop institutionalisation mechanisms that are specifically aimed at both promoting the implementation of evaluations and at improving their quality.

Mechanisms aimed at promoting the implementation of evaluations mainly consist of **providing incentives**. The most common strategy (even so, it occurs rarely) is to directly allocate public funds to carry out evaluations (Netherlands, Finland, United Kingdom). Other types of incentives may include launching national research programmes regarding public policies (Switzerland), facilitating access to data (as happens in Korea, where ministries have a legal obligation to provide and share data on public policies), organising awareness-raising meetings about evaluation (Australia) or promoting a culture of experimentation and encouraging the implementation of pilot projects (Netherlands and the United Kingdom).

Mechanisms aimed at improving the quality of evaluations are very varied. They include specific training in evaluation for public administration staff (both civil servants and, to a lesser extent, public managers and politicians), drafting guides and manuals, and establishing both evaluation quality standards and principles of ethical conduct that evaluators must follow. They also include the existence of the aforementioned

independent committees that oversee quality and/or independence of evaluations. It is striking, however, that no case study mentions the requirement that evaluations should undergo peer review, which is a practice that has been fully established in academia as a guarantee of quality.

On the other hand, the creation of communities of practice linked to evaluation and launching "what works" initiatives (or even networks of "what works" centres as in the case of the United Kingdom) are institutionalisation mechanisms that aim to both promote evaluation practices and improve its quality. Furthermore, these mechanisms also contribute to the creation of epistemic communities, which play an important role in formulating solutions to public problems (Jacob, 2023a).

In this same direction, strengthening links between public administration and external evaluators (whether they come from academia, consultancy firms or they are freelance) helps connect supply and demand, stimulating both. One manner of doing this is to create pools of evaluators that are available to public servants that commission evaluations (this is something that happens in Australia and Finland). In fact, the European Union recommends establishing permanent cooperative relationships between those who require evaluation and those who provide it (European Commission, 2008).

Table 7. Mechanisms for institutionalising evaluation: practices

<i>Institutionalisation mechanisms</i>	<i>Number of studies</i>	<i>Number of countries</i>	<i>Countries</i>
<i>Evaluation practices</i>			
<i>Evaluating more</i>			
→ Specific funds for carrying out evaluations.	3	3	FIN, NE, UK
→ National research programme on public policy.	3	1	SW
→ Access to data.	1	1	KO
<i>Evaluating better</i>			
→ Training administration staff (civil servants and politicians).	10	7	AU, CA, FIN, MX, NE, SP, SW
→ Guides and manuals.	12	8	AU, CA, CHI, FIN, KO, MX, NE, SP
→ Quality standards for evaluations.	7	5	AU, CA, MX, SW, UK
→ Principles of ethical conduct for evaluators.	3	3	AU, SW, UK
→ Peer review.	0	0	
→ Independent oversight committees to monitor quality and/or independence.	3	3	FIN, KO, NE
→ Communities of practice.	4	3	AU, CA, SW
→ "What works" initiatives.	1	1	UK
→ Evaluator pools (available to civil servants commissioning evaluations).	3	2	AU, FIN

Incorporating evaluation into decision-making processes

It is debatable whether the use of evaluations is a factor that contributes to the institutionalisation of evaluation, as proposed by Horber-Papazian and Baud-Lavigne

(2019), a dimension of it, as suggested by the institutionalisation indices presented in Table 1 and the OECD survey (2020), or whether it is actually an effect of institutionalisation itself. If we consider the use of evaluation as a dimension of institutionalisation (in a similar way to regulation, structures, and practice) it is important to analyse to what extent both the government and, above all, parliament, which is the political body that is assigned control functions, consider the results of evaluations in their decision-making processes. From the articles analysed in the scoping review, it can be inferred that in Australia, Canada, Chile, Korea, Mexico, the Netherlands, and Switzerland, there are mechanisms to ensure that the results of evaluations reach the main decision-making spaces in a regular and systematic manner so that they can be taken into account. However, only in Switzerland are studies on the actual use and influence of evaluation reports regularly published.

If we consider their use as an outcome of institutionalisation, the question is: are there mechanisms in place that are specifically aimed at establishing and improving the use of evaluations? The answer is yes. In fact, it is possible to identify mechanisms that target different points of the evaluation process. First, even before designing an evaluation, it is important to ensure that the evaluation questions accurately reflect an actual need for information. In this sense, the fact that members of parliament can commission evaluations on issues that concern them and on which they have to legislate —as happens in Korea, Finland, the Netherlands and Switzerland— is certainly an incentive to use them. Additionally, evaluators can take specific measures to ensure that their analyses are utilisation-focused (Patton, 1997).

Another strategy that contributes to the instrumental use of evaluations is to systematically bring evaluation results closer to decision-making bodies —both executive and legislative (Cordoncillo, 2023). In this sense, Goldman et al. (2023:332) identify the case of Mexico as a good practice. Here, before discussions take place on the federal budget, the evaluation body, CONEVAL, presents a document that summarises the existing evidence on national priorities and the effectiveness of existing programmes. This is in order to facilitate the use of evidence in the process of preparing and approving the budget. Another increasingly widespread mechanism is the use of response plans to address the results of evaluations. These response plans may or may not be mandatory and are aimed at specifying which recommendations will be implemented, how they will be implemented, the time frame and the person responsible. They are already common practice throughout Canada, Chile, Finland, Korea, Mexico, the Netherlands and in a few specific sectors in Spain and the United Kingdom.

Once the evaluation has been completed and delivered to the decision-making bodies, further actions can still be taken to encourage its use, even if only indirectly. Disseminating the results is key to this, and therefore the systematic publication of all the evaluations is crucial. However, this only occurs in Finland, the Netherlands, Switzerland and partially in the United Kingdom. Publishing response plans could also be considered a good practice, but only occurs in the Netherlands. Additionally, to facilitate citizens and interested parties (such as researchers or third sector entities) accessing and using these evaluations for other purposes, some countries set up evaluation repositories (Australia, Mexico and the Netherlands).

Table 8. Mechanisms for institutionalising evaluation: uses

<i>Institutionalisation mechanisms</i> Use of evaluations	<i>Number of studies</i>	<i>Number of countries</i>	<i>Countries</i>
<i>Mechanisms to promote the use of evaluations</i>			
→ Presentation of evaluation results in decision-making spaces.	8	7	AU, CA, CHI, KO, MX, NE, SW
→ Evaluation response plans.	8	6	CA, CHI, FIN, KO, MX, NE
→ Systematic publication of evaluations (and response plans).	3	3	FIN, NE, SW
→ Evaluation repositories.	3	3	AU, MX, NE
→ Studies on the use of the evaluations performed.	1	1	UK

2.5 Discussion

The mechanisms for the institutionalisation of evaluation in the political system in place in a set of 29 case studies in 12 different countries have been identified and categorised in order to develop a more comprehensive understanding of the strategies that countries can take to institutionalise policy evaluation. However, although evaluative maturity is related to the number of mechanisms deployed and most of them could be used in a complementary manner, it is clear that the process involves selecting the most appropriate ones for each context. Thus, a series of questions are suggested below. **Without being exhaustive, they could help public managers in charge of ensuring a regular use of evaluation to build favourable institutional environments and promote appropriate mechanisms.**

The first thing they should ask themselves and always keep in mind is: **what is the purpose of evaluation** (in their specific context)? That is, what is the ultimate purpose of evaluation in their political system. Our scoping review allows us to observe that evaluation pursues different objectives in the different countries analysed: while in the Netherlands it is encouraged principally for control purposes (Jacob, 2005a), in Chile the evaluation systems are clearly oriented to inform budgetary decisions "in terms of public spending and investment" (Olavarría Manríquez & Peroni Fiscarelli, 2022:196) and in Korea evaluation is closely linked to performance management. After exhaustively reviewing the literature on the purposes of evaluations, Kupiec et al. (2023) condense the answers into two types of objectives: accountability (outward-oriented) and learning (inward-oriented). They also consider these purposes as opposed to one another and mutually exclusive, since reinforcing accountability requires strengthening the norms and procedures that hinder learning (Raimondo, 2018). In this regard, it is noteworthy that countries belonging to the third and fourth waves of evaluation—likely influenced by the principles of New Public Management—tend to establish a more direct connection between evaluation and control, indicators, performance, and results.

The second issue that public managers should consider concerns the **ecosystem of agents of evaluation. Who requests and who offers evaluation** (in their specific context)? Our scoping review allows us to intuit that, while the agents that offer

evaluation do not change excessively between contexts (with some nuances that we will point out later), **the way in which the demand is organised does determine the structure of the evaluation system**. This is why, in order to understand the structure of a national evaluation system, several authors recommend considering its level of pluralism (Jacob, 2005b) or centralisation (Kupiec et al., 2023) that is, the **number of agents involved in the demand for evaluation**.

Jacob (2005b) identifies three typical scenarios: monopolies, centralised pluralism and fragmented competition. He considers monopolies to be very rare, except in environments where institutionalisation is still very weak; because “as the evaluation practice develops, a multiplication of organisations and norms aimed at the dissemination of this practice is expected” (Jacob, 2005b:56). More common is centralised pluralism: “when an organization or a norm occupies a preponderant place that determines the behaviour of the other agents” in the national evaluation system. This is the case in Romania or the Netherlands, where a unit responsible for evaluation clearly occupies a central position and the other actors have very little capacity to influence the institutionalisation of the evaluation (Jacob, 2005a; Kupiec et al., 2023). The cases of Switzerland and the United Kingdom, on the other hand, are excellent examples of the model of fragmented competence, where different types of organisations request and offer evaluations, and play a key role in promoting the quality and frequency of evaluation practice, without any one agent directing or coordinating the others.

In these decentralised models, in many cases, it is also true that **parliaments play an important role in the demand for evaluation**, a factor that some authors associate not only with evaluative maturity and greater use of evidence (Speer et al., 2015), but also with better accountability (Bundi, 2016) and efficiency in government action (Filgueiras & Queiroz, 2021). However, there are also cases with less plural and more centralised evaluation ecosystems, such as Korea, where parliamentarians also have the possibility of demanding evaluations; in this way it can be seen that they are two aspects that, although they are related, can be considered separately.

The key is not only the number and type of agents that are part of the evaluation system but also the existence of **coordination mechanisms**, which can take the shape of national or sectoral regulations, structures or strategies: what in the previous section we call the evaluation policy framework. In this way, Mexico or Canada (both federal states) represent good examples of systems with a wide plurality of well-coordinated actors. Lahey (2023) describes them more specifically as a combination of centralised regulation and supervision and decentralised delegated implementation.

How evaluation is carried out is largely determined by who is in charge of doing the evaluating. This would be the third important aspect for public managers to consider. When defining who has the competence to evaluate public policy it is important to take into account: (1) their level of independence and (2) their capacity to evaluate. Numerous authors have defended the **importance of the independence of evaluation structures**. Back in the 1990s, Weiss (1993) already pointed out that evaluations carried out by governmental actors run the risk of being less critical than those generated by non-governmental actors. This is independent of whether these external evaluations are carried out by independent bodies or units/agencies that are dependent on the legislative

branch. And Vedung (1997) also affirmed that in order to function correctly, evaluation systems need political and institutional autonomy. However, the results of our scoping review show that **in most countries the main evaluation structures depend on the executive branch**. The reasons for this generally have to do, first, with the reluctance of governments to delegate power and decision-making capacity to units that are not under their control (Filgueiras and Queiroz, 2021, citing McCubbins et al, 1987). And second, it has to do with the risk that the results of the evaluations conducted by critical and independent bodies are not taken into account. Thus, **a tension arises between the importance of protecting evaluation bodies from political pressures and a need for them to be an integral part of larger organisational structures**, so that their contributions are taken into consideration (Chelimsky, 2009).

The other structural aspect that significantly influences how evaluations are conducted—and which therefore warrants careful consideration by public managers—is the **internal capacity of the administration in question to carry out evaluations**. For governments to undertake evaluations internally, they must have access to trained and specialised personnel. As evidenced by the scoping review, such professionals may be concentrated within a specific evaluation unit or agency or distributed transversally throughout the public administration, or a combination of both. The findings also suggest that, although evaluation training for public servants is a fairly widespread strategy, few countries provide such training in a systematic and cross-cutting manner throughout the entire civil service. In this regard, Lahey (2023) distinguishes between what he calls the American model—characterized by evaluation training led primarily by academics in the social sciences—and the Canadian model, which is led from within the public administration, and involves professionals from a wide range of disciplines. Interestingly, countries with a longer tradition of evaluation—those belonging to the first and second waves—seem to have a more widespread presence of evaluation professionals across public institutions.

In that sense, it is important to keep in mind that the ability to evaluate is not only related to the practitioners' degree of knowledge about evaluation, but also to the level of formal recognition of this specialisation within the public administration. That is, to **the extent to which evaluation is professionalised** and therefore carried out by people who have the specific skills and are exclusively dedicated to evaluation tasks. The most obvious or “extreme” professionalisation strategy is the formal recognition of the evaluator figure within the public administration’s list of professional figures. However, it has already been mentioned that among the countries analysed, this has only been implemented in Canada, Switzerland and partially in Australia (three countries with a long and consolidated tradition of evaluation). Instead, in general, the results appear to point to a low level of professionalisation, despite the fact that the specialised literature indicates that “a strong profession can influence the institutional structure in which evaluators operate” (Jacob, 2023:199, referring to Cronbach et al, 1980).

In this regard, it is worth mentioning that Stockmann and Meyer (2020:188), referring to evaluation, link the issue of professionalisation to the existence of a “regular dialogue among people interested in the subject, which enables them to develop a common understanding”. This refers to the existence of **communities of practice** linked to evaluation, which according to the results of the scoping review, is a type of

institutionalisation mechanism that is rare. On the other hand, Hanwright and Makinson (2008:22) point out that the “acquisition and building of internal evaluation skills and ways of thinking will have a longer term positive impact [...] as opposed to using external evaluation sources”. This is because internal evaluation capacity translates into processes of reflection and learning that usually lead to improvements in the service or policy.

The relationship between internal evaluation capacity and evaluation **outsourcing** is not clear. On the one hand, it could be assumed that public administrations lacking the internal capacity to conduct evaluations themselves will tend to commission them to external agents —usually specialised consultancies and/or research teams. This appears to be the case in Romania and Chile. On the other hand, public administrations may choose to outsource evaluations despite having internal capacity, in order to ensure greater independence. This is the case in Germany, where external evaluations are seen as more legitimate due to their independent nature (Stockmann & Meyer, 2020). Additionally, internal evaluations generally aim to inform decisions related to the operational management of programmes, while external evaluations tend to have a more strategic focus (Stockmann & Meyer, 2020).

At the end of the day, the mechanisms that are deployed to define what an evaluation is for (to hold a body accountable or to learn and improve), what is being evaluated (what type of policies and how they are prioritised), who performs the evaluation (what type of structures) and how it is done (with what capacity and degree of independence) become relevant when the results of these evaluations are taken into account in decision-making processes, i.e. when these results are used. However, after comparing the strategies and mechanisms for institutionalising evaluation in European and American countries, Stockmann and Meyer (2022) conclude that, **although regulation, frameworks and structures are essential for embedding evaluation in the political system, they are not sufficient to guarantee it being used** (Stockmann & Meyer, 2022a). For this reason, it is important to deploy mechanisms that are specifically aimed at promoting the use of evaluations, as suggested by the categorisation proposed here.

Following the same logic, we are forced to ask ourselves whether regulation, frameworks and evaluation structures are sufficient to consolidate a culture of evaluation? Is there a point when evaluation is institutionalised enough to “guarantee” an evaluation culture? Can institutionalisation and an evaluation culture actually evolve in opposite directions? This analysis does not allow us to answer these questions, which would be interesting to investigate in future research. However, it is striking that the scoping review did not allow us to identify **any mechanism that is specifically oriented to the promotion of the values that are usually associated with an evaluation culture**. None of the identified mechanisms specifically aimed at fostering almost any of the values that, according to Mihalache (2010) (taking inspiration from Trochim, 2006), portray an “ideal-typical evaluation culture”: action-oriented, learning oriented, inclusive and participatory, responsive and fundamentally non-hierarchical, oriented towards diversity and innovation, scientifically rigorous, interdisciplinary, self-critical, honest and impartial, ethic and democratic, forward-looking and transparent. Values that, following the author, should be shared by evaluation demanders and suppliers in order to be able to talk about a national evaluation culture.

This fact is particularly concerning when various authors warn of a **risk of bureaucratisation of evaluation**. Speaking about Switzerland—a country with one of the highest levels of institutionalisation of evaluation— Jacob et al. (2015b) state that “evaluation is increasingly influenced by legalistic ways of thinking and business-influenced traditions, and the connections to a more critical social science are increasingly lost”. As a consequence, they argue that evaluation is becoming increasingly entrenched as a routine feature of bureaucracy, thereby moving away from creativity and innovation (two of the values listed above). This is a trend that Stockmann and Meyer (2022) also warn is happening globally.

The institutionalisation of evaluation cannot therefore be understood in a linear way, as a universal process (with a common scale and goals) for developing and accumulate regulations, structures and good practices. It is necessary to think strategically about how to identify and select these mechanisms to ensure that a coherent and efficient evaluation system can be constructed. But it is also necessary for this strategic reflection—as well as for the design and deployment of each of these institutionalisation mechanisms—to **take into account the values that make up the evaluation culture in each place**. These values shape the behaviour of the agents that request and offer evaluations (Mihalache, 2010) and that build the “shared epistemology” that, according to Leeuw and Furubo (2008), characterises each evaluation system. However, if these values are not promoted through specific institutional mechanisms and consequently become lost, then even though processes, structures and initiatives are implemented, evaluation runs the risk of becoming a “technology of bureaucracy” (Widmer, 2020).

2.6 Conclusion

The embedment of evaluation in a political system is related to how effective it is in terms of public policies and accountability (Figueiras & Queiroz, 2021). In a context of the rise of both populist governments and rampant disinformation, protecting evaluation with a solid legal framework and embedding it in public institutions might be essential to ensure that public policies are informed by evidence and that evaluation practices are not subject to political influences. The institutionalisation of evaluation is thus more important than ever.

This article employs a scoping review to identify and categorise the mechanisms used in 12 countries—with different evaluation traditions—to institutionalise evaluation. The analysis covers 29 studies and aims to support decision-making processes related to the institutionalisation of evaluation by offering a comprehensive, systematically organised list of institutionalisation mechanisms, as well as concrete criteria for their selection. Since the mechanisms have been inferred from the 29 studies reviewed, the list is exhaustive as a general whole, but not for each one of the countries. The categorisation includes the whole range of mechanisms that can be promoted in terms of evaluation regulation, structures, frameworks and practices, as well as a series of mechanisms aimed at promoting the use of the results of these evaluations.

Additionally, drawing on the theoretical and comparative articles identified in the review but not actually included in the scoping review—since they are not case studies—a

series of factors has been identified that public sector managers should consider when developing national evaluation systems. To build institutional environments that favour the practice of evaluation they need to reflect on why evaluation is important and what is it for (for accountability or for learning purposes), what is being evaluated (what types of policies and how they are prioritised), who is doing the evaluating (what type of structures request and offer evaluation), how it is done (with what capacity and degree of independence) and what mechanisms are available to guarantee that these results are taken into account in decision-making processes. However, the institutionalisation of evaluation is certainly necessary, but it is not sufficient to guarantee the use of evaluation (Stockmann & Meyer, 2022a), nor to consolidate an evaluation culture. In this sense, it is especially worrying that none of the mechanisms identified in the scoping review are aimed at promoting the values that are usually associated with the culture of evaluation. Without solid values, institutionalisation runs the risk of becoming mere bureaucracy. Therefore, it is urgent and necessary to identify mechanisms that are specifically aimed at promoting and consolidating the core values related to evaluation, and to analyse how they can become an integral part of institutionalisation mechanisms. This is to ensure that consolidating evaluation practices is neither an end in itself nor a purely procedural achievement, but a means to reinforce the role of evidence in political decision-making, to innovate in the design and implementation of public policies and to improve governmental effectiveness and accountability of public policies.

2.7 References

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

Alkin, M. C., Daillak, R. H., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Sage. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/using-evaluations-does-evaluation-make-difference>

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (2024). *JBI Manual for Evidence Synthesis*. JBI. <https://synthesismanual.jbi.global>

Bohni Nielsen, S., Lemire, S., & Skov, M. (2011). Measuring Evaluation Capacity—Results and Implications of a Danish Study. *American Journal of Evaluation - AM J EVAL*, 32, 324–344. <https://doi.org/10.1177/1098214010396075>

Bourgeois, I., & Cousins, J. B. (2013). Understanding Dimensions of Organizational Evaluation Capacity. *American Journal of Evaluation*, 34(3), 299–319. <https://doi.org/10.1177/1098214013477235>

Bundi, P. (2016). What Do We Know About the Demand for Evaluation? Insights From the Parliamentary Arena. *American Journal of Evaluation*, 37(4), 522–541. <https://doi.org/10.1177/1098214015621788>

Cardozo Brum, M. (2009). La institucionalización de una cultura de la evaluación en la administración pública mexicana: Avances y desafíos pendientes. *Convergencia*, 16(49), 175–198.

Chelimsky, E. (2009). Integrating evaluation units into the political environment of government: The role of evaluation policy. *New Directions for Evaluation*, 2009(123), 51–66. <https://doi.org/10.1002/ev.305>

Compton, D. W., Baizerman, M., & Stockdill, S. H. (2002). The Art, Craft, and Science of Evaluation Capacity Building. *New Directions for Evaluation*, 93, 47–61.

Cordoncillo, C. (2023). El uso de la evaluación en un contexto de institucionalización débil: Una aproximación desde el “Qualitative Comparative Analysis.” *Revista española de ciencia política*, 62, 95–119.

Dahler-Larsen, P., & Boodhoo, A. (2019). Evaluation culture and good governance: Is there a link? *Evaluation*, 25(3), 277–293. <https://doi.org/10.1177/1356389018819110>

European Union (European Commission). (2008). *Evalsed, the resource for the Evaluation of Socio-Economic Development*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2776/8296>

Filgueiras, F., & Queiroz, L. D. F. N. (2021). The Governance of Public Policy Evaluation Systems: Policy Effectiveness and Accountability. *Organizações & Sociedade*, 28(96), 208–232. <https://doi.org/10.1590/1984-92302021v28n9609en>

Furubo, J.-E., Rist, R., & Sandahl, R. (2002). *International Atlas of Evaluation*. Transaction Publishers.

Gaarder, M., & Briceno, B. (2010). Institutionalisation of Government Evaluation: Balancing Trade-offs. *The Journal of Development Effectiveness*, 2, 289–309. <https://doi.org/10.1080/19439342.2010.505027>

Goldman, I., Navarrete, T. de la G., Kalugampitiya, A., Dominguez, A. M. de E., Agbodjan, E. D., Chirau, T., & Dlakavu, A. (2023). The emergence of evaluation systems in low- and middle-income countries. In *Handbook of Public Policy Evaluation* (pp. 317–336). Edward Elgar Publishing. <http://www.elgaronline.com/edcollchap/book/9781800884892/book-part-9781800884892-29.xml>

Hanwright, J., & Makinson, S. (2008). Promoting evaluation culture: The development and implementation of an evaluation strategy in the Queensland Department of Education, Training and the Arts. *Evaluation Journal of Australasia*, 8(1), 20–25. <https://doi.org/10.1177/1035719X0800800104>

Horber-Papazian, K., & Baud-Lavigne, M. (2019). *Factors Contributing to the Strong Institutionalization of Policy Evaluation in Switzerland: Making the State Work Successfully* (pp. 355–371). https://doi.org/10.1007/978-3-319-92381-9_21

Jacob, S. (2005a). La volonté des acteurs et le poids des structures dans l’institutionnalisation de l’évaluation des politiques publiques. (France, Belgique, Suisse et Pays-Bas). *Revue française de science politique*, 55(5–6), 835–864. <https://doi.org/10.3917/rfsp.555.0835>

Jacob, S. (2005b). Réflexions autour d'une typologie des dispositifs institutionnels d'évaluation. *Canadian Journal of Program Evaluation*, 20(2), 49–68. <https://doi.org/10.3138/cjpe.20.003>

Jacob, S. (2023a). The institutionalization of evaluation around the globe: Understanding the main drivers and effects over the past decades. In *Handbook of Public Policy Evaluation* (pp. 187–205). Edward Elgar Publishing. <http://www.elgaronline.com/edcollchap/book/9781800884892/book-part-9781800884892-21.xml>

Jacob, S. (2023b). The institutionalization of evaluation around the globe: Understanding the main drivers and effects over the past decades. In *Handbook of Public Policy Evaluation* (pp. 187–205). Edward Elgar Publishing. <http://www.elgaronline.com/edcollchap-oa/book/9781800884892/book-part-9781800884892-21.xml>

Jacob, S., & Boisvert, Y. (2010). To Be or Not to Be a Profession: Pros, Cons and Challenges for Evaluation. *Evaluation*, 16, 349–369. <https://doi.org/10.1177/1356389010380001>

Jacob, S., Speer, S., & Furubo, J.-E. (2015a). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21(1), 6–31. <https://doi.org/10.1177/1356389014564248>

Jacob, S., Speer, S., & Furubo, J.-E. (2015b). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21, 6–31.

Kim, P. (2002). *Cultural Change in Government: Promoting a High-Performance Culture*. https://www.academia.edu/85704303/Cultural_Change_in_Government_Promoting_a_High_Performance_Culture

Kupiec, T., Wojtowicz, D., & Olejniczak, K. (2023). Structures et fonctions des systèmes d'évaluation complexes: Comparaison de six pays d'Europe centrale et orientale. *Revue Internationale des Sciences Administratives*, 89(1), 157–176. <https://doi.org/10.3917/risa.891.0157>

Labin, S., Duffy, J., Meyers, D., Wandersman, A., & Lesesne, C. (2012). A Research Synthesis of the Evaluation Capacity Building Literature. *American Journal of Evaluation*, 33, 307–338. <https://doi.org/10.1177/1098214011434608>

Lahey, R. (2023). John Mayne and the Origins of Evaluation in the Public Sector in Canada: A Shaping of Both Evaluation and the Evaluator. *Canadian Journal of Program Evaluation*, 37(3), 340–354. <https://doi.org/10.3138/cjpe.75451>

Lázaro, B. (2015). *Estudio comparado sobre institucionalización de la evaluación en Europa y en América Latina* (Estudios 15; Estados de La Cuestión). Programa EUROsociAL.

Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation Systems: What Are They and Why Study Them? *Evaluation*, 14(2), 157–169. <https://doi.org/10.1177/1356389007087537>

Martinaitis, Ž., Christenko, A., & Kraučūnienė, L. (2019). Evaluation systems: How do they frame, generate and use evidence? *Evaluation*, 25(1), 46–61. <https://doi.org/10.1177/1356389018802135>

McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching People to Fish? Building the Evaluation Capability of Public Sector Organizations. *Evaluation*, 9–29. <https://doi.org/10.1177/1356389003009001002>

Meyer, W., Stockmann, R., & Taube, L. (2020). The Institutionalisation of Evaluation Theoretical Background, Analytical Concept and Methods. In R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 3–34). https://doi.org/10.1007/978-3-030-32284-7_1

Mihalache, R. (2010). A Developing Evaluation Culture in Romania: Myths, Gaps and Triggers. *Evaluation*, 16(3), 323–332. <https://doi.org/10.1177/1356389010373019>

Milstein, B., & Cotton, D. (2000). *Defining concepts for the presidential strand on building evaluation capacity*. Proceedings from American Evaluation Association Conference.

OECD. (2020). *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/governance/improving-governance-with-policy-evaluation_89b1577d-en

Olavarria Manríquez, C., & Peroni Fiscarelli, A. (2022). Evaluation in Chile. In R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in the Americas* (pp. 171–199). Springer International Publishing. https://doi.org/10.1007/978-3-030-81139-6_6

Owen, J. M., & McDonald, D. E. (1999). Creating an Evaluation Culture in International Development Cooperation Agencies. *Journal of International Cooperation in Education*, 2(2), 41–53.

Patton, M. (1988a). How primary is your identity as an evaluator? *Evaluation Practice*, 9(2), 87–92.

Patton, M. (1988b). The evaluator's responsibility for utilization. *Evaluation Practice*, 9(2), 5–24.

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

Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., & Blyth, D. A. (1977). In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research. In *Using social research in public policy making* (Weiss, C.H., pp. 141–164). Lexington Books. <https://eric.ed.gov/?id=ED135938>

Pattyn, V. (2014). Explaining Variance in Policy Evaluation Regularity. The Case of the Flemish Public Sector. *Public Management Review*, 17(10), 1475–1495.

Picciotto, R. (2015). *The 5th Wave: Social Impact Evaluation*. <https://www.rockefellerfoundation.org/report/5th-wave-social-impact-evaluation/>

Preskill, H., & Boyle, S. (2008). A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation*, 29. <https://doi.org/10.1177/1098214008324182>

Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26–41. <https://doi.org/10.1177/1356389017749068>

Schein, E. H. (2010). *Organizational Culture and Leadership*. Jossey-Bass.

Segalovičienė, I. (2012). Evaluation Capacity Building in the Context of Results-Based Management. *Public Policy And Administration*, 11. <https://doi.org/10.5755/j01.ppaa.11.3.2500>

Shadish, W. R., & Luellen, J. K. (2005). History of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 183–186). Sage.

Speer, S., Pattyn, V., & De Peuter, B. (2015). The growing role of evaluation in parliaments: Holding governments accountable? *International Review of Administrative Sciences*, 81(1), 37–57. <https://doi.org/10.1177/0020852314546583>

Stockdill, S., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 2002(93), 7–26. <https://doi.org/10.1002/ev.39>

Stockmann, R., & Meyer, W. (2020). Germany. In R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 167–198). Springer International Publishing. https://doi.org/10.1007/978-3-030-32284-7_7

Stockmann, R., & Meyer, W. (2022a). The Institutionalisation of Evaluation in Europe and the Americas: A Comparison. In R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in the Americas* (pp. 509–525). Springer International Publishing. https://doi.org/10.1007/978-3-030-81139-6_17

Stockmann, R., & Meyer, W. (2022b). The Institutionalisation of Evaluation in the Americas: A Synthesis. In R. Stockmann, W. Meyer, & L. Szentmarjay (Eds.), *The Institutionalisation of Evaluation in the Americas* (pp. 451–507). Springer International Publishing. https://doi.org/10.1007/978-3-030-81139-6_16

Stockmann, R., Meyer, W., & Szentmarjay, L. (Eds.). (2022). *The Institutionalisation of Evaluation in the Americas*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-81139-6>

Stockmann, R., Meyer, W., & Taube, L. (Eds.). (2020). *The Institutionalization of Evaluation in Europe* (Vol. 34). Palgrave Macmillan Cham.

Stockmann, R., Meyer, W., & Zierke, N. (Eds.). (2023). *The Institutionalisation of Evaluation in Asia-Pacific*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-36918-6>

Toulemonde, J. (1999). Incentives, Constraints and Culutre-building as Instruments for the Development of Evaluation Demand. In R. Boyle & D. Lemaire (Eds.), *Building Effective Evaluation Capacity: Lessons from Practice*. Transaction Publishers.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>

Varone, F., & Jacob, S. (2004). Institutionnalisation de l'évaluation et nouvelle gestion publique: Un état des lieux comparatif. *Revue internationale de politique comparée*, 11(2), 271–292. <https://doi.org/10.3917/ripc.112.0271>

Vedung, E. (1997). *Public Policy and Program Evaluation*.

Vedung, E. (2010). Four Waves of Evaluation Diffusion. *Evaluation*, 16, 263–277. <https://doi.org/10.1177/1356389010372452>

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

Weiss, C. H. (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, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice*, 14(1), 93–106. [https://doi.org/10.1016/0886-1633\(93\)90046-R](https://doi.org/10.1016/0886-1633(93)90046-R)

Widmer, T. (2020). Switzerland. In R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 199–225). Springer International Publishing. https://doi.org/10.1007/978-3-030-32284-7_8

2.8 Appendix

Appendix 1: Articles included in the scoping review

Authors	Title	Publication Year	Country	Evaluation wave	Type of document
Ahonen, Pertti	Aspects of the institutionalization of evaluation in Finland: Basic, agency, process and change	2015	Finland	2a	Article
Astbury, Brad; Bayley, Scott	Evaluation in Australia	2023	Australia	2a	Book chapter
Bussmann W.	The emergence of evaluation in Switzerland	2008	Switzerland	3a	Article
Bustelo, María	The Potential Role of Standards and Guidelines in the Development of an Evaluation Culture in Spain	2006	Spain	3a	Article
Bustelo, María	Spain	2020	Spain	3a	Book chapter
Cardozo, Myriam	La institucionalización de una cultura de la evaluación en la administración pública mexicana: Avances y desafíos pendientes	2009	Mexico	4a	Article
Conde Bonfil, Carola	Avances y retrocesos de la evaluación en México. La perspectiva de los evaluadores	2017	Mexico	4a	Article
Cordoncillo, Carla	The use of evaluation in a weakly institutionalized context: a Qualitative Comparative Analysis approach	2023	Spain	3a	Article
de la Fuente Moreno A.; Cinca A.N.	El marco institucional para la evaluación de políticas públicas en España	2023	Spain	3a	Article
Gauthier, Benoît ; Lahey, Robert E.; Jacob, Steve	Evaluation in Canada	2023	Canada	1a	Book chapter
Goswami S.; Lane V.	Building Evaluation Capital in Government: A Queensland Departmental Approach	2017	Australia	1a	Article
Hanwright J.; Makinson S.	Promoting evaluation culture: The development and implementation of an evaluation strategy in the Queensland Department of Education, Training and the Arts	2008	Australia	1a	Article
Horber-Papazian K.; Baud-Lavigne M.	Factors Contributing to the Strong Institutionalization of Policy Evaluation in Switzerland	2019	Switzerland	3a	Article

Horber-Papazian K.; Rosser C.	From law to reality: A critical view on the institutionalization of evaluation in the swiss canton of geneva's parliament	2018	Switzerland	3a	Book chapter
Horber-Papazian, Katia; Jacot-Descombes, Caroline	Is Evaluation Culture Shaped by the Swiss Political System and Multiculturalism?	2012	Switzerland	3a	Book chapter
Huckel Schneider C.; Milat A.J.; Moore G.	Barriers and facilitators to evaluation of health policies and programs: Policymaker and researcher perspectives	2016	Australia	2a	Article
Jacob S.	La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques (France, Belgique, Suisse et Pays-Bas)	2005	Netherlands	2a	Article
Jacob S.	La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques (France, Belgique, Suisse et Pays-Bas)	2005	Switzerland	3a	Article
Klein , Carolien	The Netherlands	2020	Netherlands	2a	Book chapter
Lahey, Robert	John Mayne and the Origins of Evaluation in the Public Sector in Canada: A Shaping of Both Evaluation and the Evaluator	2023	Canada	1a	Article
Mihalache R.	A developing evaluation culture in Romania: Myths, gaps and triggers	2010	Rumania	4a	Article
Mora C.; Antonie R.	Levers supporting program evaluation culture and capacity in Romanian public administration: The role of leadership	2012	Rumania	4a	Article
Olavarría, Claudia; Peroni, Andrea	Evaluation in Chile	2022	Chile	3a	Book chapter
Park, Nowook	Evaluation in Korea	2023	Korea	4a	Book chapter
Salvador, Janett; Meza, Jaqueline	Evaluation in México	2022	Mexico	4a	Book chapter
Stockmann R.; Meyer W.	Germany	2020	Germany	1a	Book chapter
Uusikylä, Petri; Ahonen, Pertti; Hietakangas, Nina K.; Nurmi, Johanna	Finland	2020	Finland	2a	Book chapter
Valiño, Aurelia	Evaluation in public policy analysis: the state of the art in Spain [La evaluación económica en el análisis de las políticas públicas: el estado del arte en España]	2023	Spain	3a	Article
Widmer, Thomas	Switzerland	2020	Switzerland	3a	Book chapter

Wond, Tracey	United Kingdom	2020	United Kingdom	1a	Book chapter
Jacob, Steve; Speer, Sandra; Furubo, Jan-Eric	The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later	2015	Several	NA	Article
Pattyn V.; van Voorst S.; Mastenbroek E.; Dunlop C.A.	Policy evaluation in Europe	2017	Several	NA	Book Chapter
Stockmann R.; Meyer W.	The Institutionalisation of Evaluation in the Americas: A Synthesis	2022	Several	NA	Book Chapter
Stockmann R.; Meyer W.	The Institutionalisation of Evaluation in Europe and the Americas: A Comparison	2022	Several	NA	Book Chapter
Trujillo C.M.	CLEAR LAC-Centers for Learning on Evaluation and Results-Latin America and the Caribbean	2022	Several	NA	Book Chapter
Varone F.; Jacob S.	Institutionnalisation de l'évaluation et nouvelle gestion publique: Un état des lieux comparatif	2004	Several	NA	Article
Chaytor, Kaireen	Building Evaluation Culture-The Missing Link	2023	None	NA	Article
Goldman, Ian; de la Garza, Thania; Kalugampitiya, Asela; de Erice, Alonso Miguel; Djimitri Agbodjan, Edoé; Chirau, Takunda; Dlakavu, Ayabulela	Chapter 20: The emergence of evaluation systems in low- and middle-income countries	2023	None	NA	Book Chapter
Jacob S.	Réflexions autour d'une typologie des dispositifs institutionnels d'évaluation	2005	None	NA	Article
Jacob, Steve	Chapter 12: The institutionalization of evaluation around the globe: understanding the main drivers and effects over the past decades	2023	None	NA	Book Chapter
Leeuw F.L.; Furubo J.-E.	Evaluation systems: What are they and why study them?	2008	None	NA	Article
Meyer W.; Stockmann R.; Szentmarjay L.	The Institutionalisation of Evaluation: Theoretical Background, Analytical Concept and Methods	2022	None	NA	Book Chapter
Meyer W.; Stockmann R.; Taube, L.	The Institutionalisation of Evaluation: Theoretical Background, Analytical Framework and Methodology	2023	None	NA	Book Chapter
Meyer, Wolfgang ; Stockmann, Reinhard; Taube, Lena	The Institutionalisation of Evaluation: Theoretical Background, Analytical Concept and Methods	2020	None	NA	Book Chapter

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

Autora: Carla Cordoncillo Acosta

Publicación: Cordoncillo, C. (2023). El uso de la evaluación en un contexto de institucionalización débil: una aproximación desde el qualitative comparative analysis. *Revista Española De Ciencia Política*, (62), 95–119. <https://doi.org/10.21308/recp.62.04>

Resumen

El auge de las políticas basadas en evidencias ha revivido el interés por los usos instrumentales de las evaluaciones. Sin embargo, la evidencia disponible sobre usos instrumentales 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 en otras investigaciones realizadas 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.

3.1 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.

Los usos instrumentales sí son, no obstante, el tipo de uso más estudiado. Desde mediados de los años setenta, 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, sin embargo, 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* (en adelante, 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.

3.2 Marco teórico y 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 se refieren al cierre o la continuidad del programa o la política, a su reducción o su expansión; pero también pueden implicar poner en marcha una nueva política o realizar modificaciones sustantivas a la política evaluada. 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*, los cuales refieren a los beneficios obtenidos por los *stakeholders* que se involucran en la evaluación.

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 la política pública (Rein y White, 1977) o el nivel de implicación de quienes toman las 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 persona evaluadora y <i>policymaker</i>
			Características del evaluador/a
	Factores técnicos (metodológicos)	Proceso de evaluación	Características de la administración pública (organizacionales y de quien gestiona la evaluación)
			Características de la evaluación

Fuente: elaboración propia a partir de las fuentes citadas.

A lo largo de los años noventa, 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).

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 los Países Bajos 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 desigualdades profundas en el nivel de institucionalización de la evaluación entre los países de la Organización para la Cooperación y el Desarrollo Económicos (en adelante, OCDE). Aunque algunas de estas diferencias se han matizado recientemente en un nuevo ejercicio analítico en profundidad sobre la institucionalización de la evaluación en Europa (Stockmann et al., 2020), los estudios que muestran el efecto de la cultura política nacional (Barbier y 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 continúan siendo válidos. Así, 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 años noventa (como España), ésta 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 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? 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 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: 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.

3.3 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 evaluación⁴, 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 años setenta; (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

⁴ La posición 17 deriva de la suma de las puntuaciones que se exponen a continuación para cada uno de los criterios considerados por Jacob et al. (2015): (1) La evaluación de políticas públicas tiene lugar en distintos ámbitos sectoriales: 1,3; (2) Hay oferta de personas evaluadoras provenientes de distintas disciplinas: 1,8; (3) Existe un discurso nacional relativo a la evaluación: 1,5; (4) Existen organizaciones profesionales de evaluadores/as: 2,0; (5) Grado de institucionalización de la evaluación en el Gobierno: 1,3; (6) Grado de institucionalización de la evaluación en el Parlamento: 0,5; (7) Pluralismo de quienes realizan evaluaciones en cada ámbito sectorial: 1,3; (8) Práctica de la evaluación dentro de la institución superior de auditoría: 0,3; (9) Proporción de evaluaciones de impacto y de resultados, en relación con las de proceso o implementación: 1,5. Así, España sumó un total de 11,3 puntos, sobre un total de 18.

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).

3.4 Casos de estudio

En este artículo se analizan 20 evaluaciones realizadas por el Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua) – un consorcio público constituido por el Departamento de Economía y Hacienda de la Generalitat de Catalunya, la Diputación de Barcelona y la Universidad Pompeu Fabra- 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.

Las 20 evaluaciones fueron encargadas a Ivàlua por administraciones públicas de distinta naturaleza: departamentos de la Generalitat, la Diputación de Barcelona, ayuntamientos y organismos públicos, como el Consorcio de Educación de Barcelona o Barcelona Activa. A pesar de ser algunas de ellas entidades consorciadas, Ivàlua es un organismo independiente, que trabaja en base a procedimientos estandarizados que no dependen de la relación con la organización contratante. Así, el hecho de que todas las evaluaciones fueran realizadas por la misma entidad (una agencia que además es 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. También es importante mencionar que ninguna de las políticas consideradas en el análisis es multi-nivel, lo cual evita una posible influencia de niveles institucionales superiores tanto en el uso de la evaluación como en la forma como se han desplegado las distintas condiciones. Y solamente 2 de las políticas evaluadas recibieron fondos europeos, pero -aunque dicha financiación pudo condicionar la realización de una evaluación (que es en muchos casos una contraprestación obligatoria a los fondos)- no

es de esperar que determinara el uso que se hizo de sus resultados, pues este no es un aspecto sobre el cual las instituciones europeas pidan información alguna.

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 a la persona que coordinaba cada uno de los equipos de evaluación.

3.5 Metodología

El *Qualitative Comparative Analysis* (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 equifinalidad 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.

3.5.1 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 colocado los métodos experimentales (o en su defecto los cuasiexperimentales) en la cúpula de lo que el mismo movimiento ha denominado “jerarquías de evidencias” (Parkhurst, 2017) y ha considerado la evidencia que se genera a partir de estos como la más adecuada para informar el diseño de las políticas públicas (Parkhurst, 2017). Sin embargo, este enfoque -extremadamente economicista y positivista- ha sido ampliamente contestado por diversas personas investigadoras en el ámbito de la evaluación y la inferencia causal -incluso por sociedades de evaluación como la European Evaluation Society (European Evaluation Society, 2007)-, quienes han cuestionado su cooptación de los términos “impacto” y “evidencias” y se han esforzado por poner en valor otros métodos y enfoques igualmente útiles (Befani, 2016; Befani et al., 2014; Stern et al., 2012), así como por defender un acercamiento multi-método a la evaluación de impacto. Aún así, el debate entre los llamados “randomistas” y sus detractores continúa abierto, por lo que la primera condición considerada en este estudio es si el caso corresponde a una evaluación de impacto (realizada a partir de métodos experimentales o cuasiexperimentales) o no (en adelante, condición IMP). De esta forma, se espera no solamente aportar evidencias sobre la influencia del método en el potencial de uso de la evaluación, sino también contribuir al debate.

Balthasar (2006) define los factores burocráticos o relativos al contexto institucional como aquellos que marcan la distancia entre personas evaluadoras y evaluadas. Sin embargo, los 6 factores que propone no resultan relevantes en este caso por falta de variación. Y es que 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 quienes evalúan y quienes deciden con base a los resultados en función de 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 el equipo evaluador y el *policymaker* (en adelante, 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 (en adelante, 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 (en adelante, 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 (en adelante, 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, de acuerdo con la definición proporcionada en el marco teórico. Para más información sobre la selección de factores pueden consultarse los materiales adicionales.

3.5.2 Calibraje

El QCA requiere transformar la información empírica disponible sobre cada uno de los casos en niveles de pertenencia al conjunto que representa cada una de las condiciones; este proceso se denomina “calibraje” (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 las 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 y 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 calibraje, así como la tabla con las cuatro condiciones y el resultado calibrados para las 20 evaluaciones (*raw data*).

3.6 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	Consistency	Coverage	RoN	Condición	Consistency	Coverage	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 así, de los 9 casos que derivaron en un uso instrumental en Cataluña, 6 son evaluaciones realizadas a partir de métodos experimentales o cuasi-experimentales (o evaluaciones económicas, basadas también parcialmente en este tipo de métodos).

3.6.1 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)⁵. 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.

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).

⁵ 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*.

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 policymaker	●	
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 personas evaluadoras y evaluadas 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.

3.7 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. El caso de Cataluña parece apuntar, 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 las dos soluciones del QCA, 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 por Pattyn y Bouterse (2020): 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 diseminación, formal o informal? Una posible explicación la podemos encontrar en la idea de las “dos comunidades” de Caplan (1979), según la cual la academia -o en este caso las personas evaluadoras- y quienes toman las decisiones de política pública 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 y las *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 las 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.

En cambio, la relación entre equipo evaluador y *policymaker* (que ellas identifican como constante en todos sus casos) se muestra una condición con amplia variabilidad en Cataluña. El grado de implicación de los y las *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) también aparece como relevante para explicar los usos instrumentales de la evaluación en el contexto catalán -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 y 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 Vilchez, 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-experimentales⁶. Una posible explicación de esta diferencia podría ser que, en contextos donde la práctica evaluativa es menor, tiendan a priorizarse las evaluaciones de impacto para mostrar un compromiso, más simbólico que estructural, con la evaluación y las políticas informadas en evidencias. En cambio, en contextos como los Países Bajos, con mayor cultura de la evaluación, está más asentada la aproximación multi-método.

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 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 como es el catalán, aquellas

⁶ 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 amplio abasto, con numerosos objetivos que se traducen también en diversas intervenciones, 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.

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 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 coyuntural⁷ del uso de la evidencia ya apuntada por Ledermann (2012).

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

⁷ Relativa a causalidad coyuntural (*conjunctural causation*).

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; 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 que tienen mayor credibilidad según el movimiento de políticas basadas en evidencias -que ha ganado espacio en los últimos años en el ámbito de la gestión pública-, como forma de mostrar un compromiso más simbólico que estructural de las administraciones públicas con la evaluación. 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 -según los economistas- 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.

3.9 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, Emilio, 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.

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>

Befani, Barbara (2016). *Choosing Appropriate Evaluation Methods. A Tool for Assessment and Selection*. Bond. <https://www.bond.org.uk/resources/evaluation-methods-tool>

Befani, Barbara, Barnett, Chris y Stern, Elliot (2014). Introduction. Rethinking Impact Evaluation for Development. *IDS Bulletin*, 45 (6).

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 Wolfgang Meyer, Reinhard Stockmann y Lena Taube, (ed.), *The Institutionalisation of Evaluation in Europe*. Springer International Publishing.

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>

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 Vilchez, 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>

European Evaluation Society (2007). *EES Statement: The importance of a methodologically diverse approach to impact evaluation - Specifically with respect to development aid and development interventions*. <https://europeanevaluation.org/wp-content/uploads/2020/03/EES-Statement-on-methodological-diversity.pdf>

Furubo, Jan-Eric y Sandahl, Rolf. (2002). "Introduction. A diffusion perspective on Global Developments in Evaluation", en Jan-Eric Furubo, Ray Rist y Rolf Sandahl, (ed.), *International Atlas of Evaluation*, 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.

- 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*. Bruselas: PIE Peter Lang.
- 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 Carol Weiss, (ed.), *Using Social Research in Public Policy Making*, 165-182. Lexington: 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", en Reinard Stockmann, Wolfgang Meyer y Laszlo Szentmarjay, (ed.), *The institutionalization of Evaluation in the Americas*, 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 Carol Weiss, (ed.), *Using social research in public policy making*, 141-164. Lexington: 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. y Fiss, Peer C. (2008). "Net Effects Analysis Versus Configurational Analysis: An Empirical Demonstration". En Charles Ragin, (ed.), *Redesigning Social Inquiry: Set Relations in Social Research*. 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.

Stern, Elliot, Stame, Nicoletta, Mayne, John, Forss, Kim, Davies, Rick y Befani, Barbara (2012). *Broadening the Rang of Designs and Methods for Impact Evaluation*. <https://www.oecd.org/derec/50399683.pdf>

Stockmann, Reinhard, Meyer, Wolfgang y Taube, Lena (2020). *The Institutionalization of Evaluation in Europe*. Palgrave Macmillan Cham.

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 condicions 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 (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 Evluative Activities. En Jean-Claude Barbier y Penny Hawkins, (ed.), *Evaluation Cultures: Sense-Making in Complex Times*. New Bruswick: Transaction Publishers.

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

4. Fostering innovation through collaboration: A comparison of collaborative approaches to policy design

Authors: Carla Cordoncillo Acosta and Mireia Borrell Porta

Publication: This is a post-peer-review, pre-copy edited version of an article published in *Policy & Politics*. The definitive publisher-authenticated version of Cordoncillo Acosta, C., & Borrell-Porta, M. (2025). Fostering innovation through collaboration: a comparison of collaborative approaches to policy design. *Policy & Politics* is available online at <https://doi.org/10.1332/03055736Y2024D000000058>

Abstract

Wicked problems have forced policymakers to develop new strategies for policy design that deal better with complexity. Both cross-sectoral and multi-actor collaboration are presented by collaborative governance advocates as potential solutions. Nevertheless, even if both approaches have been theoretically linked to policy innovation, there is little empirical evidence to support those relationships and none that compare the innovative potential of different collaborative approaches. Using regression analysis, we analysed 529 policy strategies promoted by the Barcelona City Council as part of the *Pla de Barris* strategy to identify whose participation in policy design is more meaningful in terms of innovation. Within the public sector, both collaborative government and cross-sectoral collaboration appear to be related with policy innovation. However, when non-public actors—in particular, third-sector organisations—get involved in the policy design process (through co-creation), the innovative potential of the policy output increases significantly.

4.1 Introduction

In a context where societal problems are characterized by being “wicked” (Geyer & Rihani, 2010; Wagenaar, 2007) and governments’ power is increasingly fragmented and distributed (Rhodes, 2007), policy innovation—defined as new objectives, instrument and assessment tools that disrupt the common wisdom and established practice in a particular context in order to provide more effective solutions (Torfing, 2016)—has become an imperative. Policymakers need to come up with new policy strategies of intervention that deal better with complexity. Additionally, citizens are not only demanding to have more voice in decision-making processes but also have raised their expectations of governments’ actions (Bentzen et al., 2020). As a consequence, even if systematically exploring new directions for better policies and services is not something public administrations usually spend their time on (Bason, 2018), public managers are forced to renew their approaches and tools for policy design (Bason, 2017; Bourgon, 2011; Head, 2022).

In their own way, both Weberian bureaucracy and New Public Management have emphasised a culture of hyper-specialisation (B. Crosby et al., 2016a; Wagenaar, 2007). Nevertheless, today's societal problems cannot be broken down into parts and solved by a single public organisation (Bason, 2010; Doz & Koskonen, 2014). Advocates of a paradigm shift towards networked or collaborative governance (Ansell & Torfing, 2014; Bingham et al., 2005; Emerson et al., 2012; Hartley, 2005; Osborne, 2006; Paquet, 2009) argue that addressing wicked problems requires collaboration among multiple actors. Incorporating diverse perspectives into the policy process is believed to stimulate creativity and generate synergistic results that would be impossible for those same actors to achieve on their own (Bason, 2010, 2017; Straus, 2002; Torfing, 2016). And that is why multi-actor collaboration is said to stimulate public innovation in general (Bason, 2010; Torfing, 2016), and particularly policy innovation.

Nevertheless, even if the relationship between collaboration and policy innovation has caught a lot of attention recently around the concept of "collaborative innovation" (B. Crosby et al., 2016b; Hartley et al., 2013; Nambisan, 2008; Sørensen & Torfing, 2011; Torfing, 2016), most of the contributions have been theoretical. Furthermore, among the few empirical ones (Bommert, 2010; Krogh & Torfing, 2015; Newman et al., 2001; Torfing, Krogh, et al., 2020), none compares the innovative potential of different collaborative arrangements —understood as different combinations of policy actors— in order to understand whose collaboration in the policy design process is more meaningful in terms of innovation. Consequently, it is difficult to offer policymakers specific advice on how to handle policy design from a collaborative approach.

In this article, we analyse the actors involved in designing 529 policy strategies promoted by the Barcelona City Council under *Pla de Barris* initiative. The goal is to determine whether some collaborative arrangements are more likely to result in policy innovation and identify which ones. Using regression analysis, we compare the innovative potential of these arrangements and complement existing qualitative evidence on collaborative innovation.

Results, first, corroborate that policy design processes involving several actors —whether public or non-public— are more likely to result in policy innovation than those involving a single specialized public actor. Both cross-sectoral collaboration and collaborative government appear to be powerful approaches to foster policy innovation within the public sector, without needing to open the policy design process to non-public policy actors. Nevertheless, when some of those non-public actors get involved, the innovative potential of the policy output increases significantly. Hence, with our research we hope to influence policymakers who are looking for new and better solutions to societal problems in the management of policy design processes.

4.2 Dealing with complexity from a collaborative approach

Since Rittel and Webber (1973) argued that technocratic approaches were no longer adequate to tackle some issues of social policy, a lot of attention has been paid to characterising societal problems and understanding this new and complex context. The proliferation of theoretical and empirical studies about "wicked problems" —understood

as ill-defined problems, that cannot be definitively solved since they are dependent on “elusive political judgment for resolution” (Rittel & Webber, 1973:160)—evidences a general concern about complexity, as well as a need for new approaches to policymaking.

Linear and standardised procedures are no longer useful to answer most societal problems (Head, 2022). The cross-cutting character of wicked problems directly challenges the simplistic sectoral approach that has characterised public sector bureaucracy (B. Crosby et al., 2016a), even after New Public Management (NPM) reforms (Wagenaar, 2007). In fact, “the strong focus of NPM on managerialism, budget discipline, performance targets and the use of conditionals rewards” seems to have strengthened administrative silos (Torfing et al., 2020:128). However, wicked problems cannot be broken down into component parts; they demand to be analysed holistically and they require cross-sectoral solutions (Agranoff, Robert, 2003; Weber & Khademian, 2008), since in complex systems, the whole exhibits properties that cannot be explained by understanding its parts separately (Kauffman, 1995 cited in Wagenaar, 2007).

Therefore, public managers and policymakers are forced to renew their approaches and tools (Bason, 2017; Bourgon, 2011; Head, 2022). Policy design—defined as the effort to develop effective policies through the application of knowledge about policy means (Howlett, 2014)—requires considering a broader range of contingencies and interrelated factors. This complexity compels decision-makers and public managers to work unitedly, rather than from disciplinary silos (Doz & Koskonen, 2014). Addressing problems comprehensively calls for a kaleidoscopic perspective (Kanter, 1988). This is why proponents of a paradigm shift in public management, toward what has been called networked governance (Hartley, 2005), collaborative governance (Ansell & Torfing, 2014; Emerson et al., 2012; Paquet, 2009), or new public governance (Bingham et al., 2005; Osborne, 2006), argue that past reforms failed to translate organisational and procedural innovations into policy innovation (Fung & Wright, 2003; Torfing, 2016).

Innovation is defined as the development of “new and creative ideas that disrupt the common wisdom and established practice in a particular context in order to provide more effective solutions” (Torfing et al., 2020:399, citing Torfing, 2016). However, when focusing on the public sector, Torfing (2016) distinguishes between product, service, process, organisational, governance, policy and discourse innovations. A seven-categories typology that he proposes to collapse into three broader ones, since in practice they tend to overlap: service innovation, organisational innovation, and policy innovation, defined as “new objectives, instruments, and assessment tools” (2016:37). Thus, following Torfing (2016), we define policy innovation as new objectives, instrument and assessment tools that disrupt the common wisdom and established practice in a particular context in order to provide more effective solutions. A definition that, similarly to that of Paz and Fontaine (2018:3)—“the introduction of disruptive, original, hitherto unseen and disturbing practices that permanently alter the fundamentals of a policy by moving away from its core *status quo ante*—emphasizes three aspects: (1) novelty, (2) in a specific context, (3) concerning some of the components of a public policy. A change that, though it must be implemented to take effect, first must occur during the policy design phase. If the ultimate goal is to handle complexity better, innovation (disruption) needs to happen when the problem is being framed and the solution is proposed.

In that direction, if the fragmented specialisation culture is perceived by collaborative governance researchers as one of the main barriers to policy innovation, both cross-sectoral and multi-actor collaboration are presented as potential solutions. In fact, they are not mutually exclusive, and they both imply collaboration, defined as “a temporal process through which a plurality of actors work together in an organized way to transform problems and opportunities into joint solutions that rest on provisional agreements that are formed despite the persistence of various forms of dissent” (Torfing, 2016:64). Cross-sectoral approaches focus on collaboration among public organisations (such as departments, agencies, and units) particularly in problem-framing. In contrast, multi-actor collaboration emphasises the participation of non-public actors (citizens, third sector organisations, academia, and private companies), which can result in a wide range of combinations between public and non-public actors involved in policy design.

Nevertheless, even if both approaches have been theoretically linked to policy innovation (Bason, 2018; Torfing, 2016), there is few empirical evidence sustaining that relationship and none comparing their innovative potential. Is intersectorality *enough* for policy innovation? Are policy design processes involving non-public actors always more likely to result in policy innovation than those managed by public organisations? Without specific comparative information on the innovative potential of those different collaborative arrangements it is difficult to offer policymakers specific advice on how to handle policy design processes from a collaborative approach and how to foster policy innovation.

4.2.1 Empirical evidence

Most of the literature defending a causal relationship between collaboration and innovation dates from the 90s (Borins 1998; Roberts and King 1996; Roberts and Bradley 1991; Van de Ven et al. 2008), when collaborative governance was not even a concept. Furthermore, with the exception of Roberts and Bradley' (1991) qualitative longitudinal study, none of those researches focused on policy innovation. For a long time, the attention was oriented towards public sector's capacity for innovation (its determinants, barriers, risks and facilitators) and, thus, towards organizational innovation (Ansell & Torfing, 2014; Arbolino et al., 2019; Arundel et al., 2019; Bason, 2010; Bhatta, 2003; Borins, 2001; Daglio et al., 2014; Damanpour, 1991b; Eggers & Singh, 2009; Mulgan, 2007; Mulgan & Albury, 2003; Van de Ven et al., 2008).

Some years later, the notion of “collaborative innovation” —defined as a “collaborative approach to innovation and problem solving in the public sector that relies on harnessing the resources and creativity of external networks and communities” (Nambisan, 2008:11)— synthesized the academics and practitioners' interest in enhancing public innovation through collaboration. However, on the one hand, its literature remains mostly theoretical and descriptive (B. Crosby et al., 2016b; Hartley et al., 2013; Nambisan, 2008; Sørensen & Torfing, 2011; Torfing, 2016) and the few existing empirical analysis are based on case studies (Bommert, 2010; Krogh & Torfing, 2015; Newman et al., 2001; Torfing, Krogh, et al., 2020). On the other hand, if it has already been stated that “it is difficult to separate the process of collaboration from the innovative results and their problem-solving effects” (Torfing et al., 2020:401, citing Innes & Booher, 1999) the notion of “collaborative innovation” makes it even harder.

That partially explains why those analysis tend to treat collaborative approaches “as a whole”, at least on what concerns its link with policy innovation. Even if some publications distinguish different kinds of institutional collaborative arrangements and theoretically link them to policy innovation (Eggers & Singh, 2009; Hartley et al., 2013), none compare their innovative potential. Therefore, there is a need for empirical evidence considering the diversity of public and non-public agents’ arrangements behind the category of “collaborative approaches” and analysing the innovative potential of the different combinations.

4.2.2 Research question and hypotheses

As exposed, theoretical literature clearly explains why today’s societal problems cannot be solved through the isolated efforts of a single authority (Bason, 2018; B. Crosby et al., 2016a) and, hence, require collaboration. It is the integration of different perspectives what stimulates creativity and produces synergistic results that would be impossible for those same actors to achieve on their own (Bason, 2010, 2017; Straus, 2002; Torfing, 2016). Furthermore, the constructive management of differences brings on a better understanding of the problem at hand, enhances mutual learning and facilitates the generation of new ideas that, based on a broad set of experiences, disturb each other’s perceptions of the world (Torfing, Krogh, et al., 2020). That is why we expect collaborative approaches to policy design to be more likely to result in policy innovation than traditional specialized ones.

However, since collaboration can take place in extremely different conditions among very different actors (Torfing, 2016), in order to provide specific recommendations for policymakers we need to understand whose participation in the policy design is more meaningful in terms of innovation. We need to analyse **whether some collaborative arrangements have more innovative potential than others and identify which ones**. Based on the mentioned theoretical literature, we propose various hypotheses about which arrangements are more likely to lead to policy innovation, helping to identify which policy actors should get involved in the policy design process.

First, though, we verify that policy design processes involving several actors (either public or non-public) are indeed more likely to result in policy innovation than policy design processes involving a single public actor (**H1**). And make sure that we obtain the same result when only public actors are involved in the design process, since the literature is clear when stating that contemporary policymaking surpasses formal organisational boundaries (Jones et al., 1997; Shearer et al., 2016; Sørensen & Torfing, 2009). Namely, we verify that policy design processes involving several public actors are more likely to result in policy innovation than policy design processes involving a single public actor (**H2**). According to Bason (2018:109), “the actual policy or service design must inevitably be the result of multiple departments, agencies and other actors working closely together in new ways, achieving their results through others, not only on their own”.

However, what has been called “joined-up”, “collaborative” or “networked” government (Bason, 2010; Eggers & Singh, 2009; Mulgan, 2009) —that is to say, different forms of collaboration among public organisations— can take too many shapes: collaboration may take place between very different public actors. Thus, to provide specific

recommendations we need to understand who must collaborate with who. Theoretical literature clearly points to hyper-specialisation culture as a barrier for both solving complexity management and fostering policy innovation (Bason, 2018; Eggers & O'Leary, 2009; Eggers & Singh, 2009). On the one hand, an integral understanding of problems requires a kaleidoscopic vision (Kanter, 1988), which can hardly come from a single disciplinary silo. On the other hand, innovation is positively correlated with specialization diversity (Damanpour, 1991b; Kimberly & Evanisko, 1981). Therefore, we expect policy design processes involving public actors from different disciplinary silos to be more innovative than those approached from a sectoral perspective. Furthermore, sectorial collaboration could easily fit into what NPM defined as corporate management approach—consisting on putting “an executive director in charge of several related silos in order to promote intra-organizational coordination” (Torfing et al., 2020:131)—and it has already been stated that NPM failed to address wicked problems (B. Crosby et al., 2016a; Wagenaar, 2007). Namely:

H3: Cross-sectoral policy design processes are more likely to result in policy innovation than sectoral ones. In other words: policy design processes involving several public actors representing different disciplinary silos are more likely to result in policy innovation than policy design processes undertaken by different public actors from the same silo.

Still, many authors claim that public sector innovation requires collaborative interaction between public and private actors (Bommert, 2010; Borins, 2001; Eggers & Singh, 2009; Nambisan, 2008). Even if there is no empirical evidence comparing the innovative potential of only-public and public-private networks, a high diversity of actors is positively associated with radical innovation (Dente et al., 2012; Howlett, 2002), and it is plausible to expect that the divergence of interests, experiences, ideas, skills and resources will be higher between public and private actors than among the firsts. In fact, according to Torfing et al. (2020:402), “since public employees often think along similar lines, it is even better [to stimulate out-of-the-box thinking] if private for-profit or non-profit organisations are included in the collaborative arena”. As a consequence, we hypothesise that:

H4: Policy design processes involving non-public actors are more likely to result in policy innovation than policy design processes undertaken by several public actors.

According to co-creation researchers, however, it is the participation of users or citizens in the policymaking what has the potential to generate new policy solutions to complex societal problems (Ansell & Torfing, 2021; Bentzen et al., 2020; B. Crosby et al., 2016a; B. C. Crosby & Bryson, 2010; Torfing et al., 2021). Even so, despite being insistently claimed, the innovative potential of co-creation—understood as “a joint effort of citizens and public sector professionals in the initiation, planning, design and implementation of public services” (Brandsen et al., 2018:3)—has little empirical support, since much of its research remains theoretical and descriptive (Verschuere et al., 2012). Still, based on the few existing case studies, which point to a positive relationship between co-creation and policy innovation (Brandsen et al., 2018; Nabatchi et al., 2017), and taking into account the also positive relation between diversity and innovation (Dente et al., 2012; Howlett, 2002) we hypothesise that:

H5: Policy design processes involving citizens are more likely to result in policy innovation than policy design processes in which they are not involved.

Moreover, we have decided to explore the innovative potential of other types of non-public actors, such as third-sector organisations, neighbourhood associations and private companies, with the aim of understanding how policymakers can spur policy innovation by inviting the most meaningful actors to take part in policy design processes.

We propose a regression analysis to compare the innovative potential of those different collaborative approaches. Thereby, we not only provide empirical evidence sustaining the relation between networked government, cross-sectoral collaboration and public – non-public collaboration with policy innovation, but we also identify which collaborative arrangements are more meaningful—in terms of innovation—and complement existing evidence on the relation between collaboration and policy innovation with a quantitative analysis, which are scarce.

4.3 Case selection and data operationalisation

To empirically test our hypotheses and understand the way the different collaborative arrangements relate to policy innovation, we use the programme *Pla de Barris* as a case study. Promoted by the Barcelona City Council, *Pla de Barris* is a macro-strategy against urban segregation that brings together hundreds of public policy interventions in very different policy domains—education, sustainability, public space, social economy, housing, employment, community action, gender equality, among others—to be implemented in the 10 poorest neighbourhoods of Barcelona. Some of those policy interventions were designed and directly promoted by the City Council and others were the result of bottom-up processes and collaborations with many other policy actors (other public entities, third-sector organisations—which are service-oriented entities that are neither public nor private—, neighbourhood associations, non-organised citizens and private companies). Thereby, acting mostly as a 150-million-euro fund to spend in 4 years (2016-2020), the strategy ended up covering 529 policy interventions⁸.

What is meaningful for the purposes of this research is that some of those 529 policy interventions were designed by a single unit (department, area or agency) of the City Council and others were the result of a collaborative arrangement among either public agents (only) or public agents together with different non-state actors. Indeed, this case is a good example of the fact that, even within a public administration that has a consolidated participatory tradition and is strongly committed to collaborative governance (Parés et al., 2015), collaborative policy design initiatives coexist with traditional policy design processes (Torfing, 2016).

The analysis is based on the monitoring data collected by a team of project managers over the 529 policy interventions funded by *Pla de Barris*. Thus, there was no selection.

⁸ *Pla de Barris*' budget ended up covering 529 social policy interventions -policy programs and actions of different policy domains- and 193 public work interventions. However, since public works have a different nature, more associated to architectural parameters, and it is difficult to link them to policy innovation, they have not been considered in the analysis.

Since our interest lies in the design phase, we considered all the interventions that were ready to start its implementation phase, whether the implementation was successful or not. For each one we know which actors participated in the design process, and we have a detailed description of the resulting policy intervention, as well as other features, such as its domain, timings and budget. From the list of participating actors, we can infer the type of collaborative arrangement. Therefore, independent variables were operationalised as follows:

RQ1–H1: *collaborative approach (two or more policy actors involved, either public or not) vs. a single public actor involved*

RQ2–H2: *collaborative government (several public actors involved) vs. a single public actor involved*

RQ2–H3: *cross-sectoral approach (several public actors from different policy domains involved) vs. sectoral approach (several public actors from a single policy domain involved)*

RQ2–H4: *collaborative government (several public actors involved) vs. co-creation (public and non-public actors involved)*

RQ2–H5: *co-creation with each type of non-public actor separately*

Operationalising the **dependent variable** (policy innovation) is not an easy task, since it is difficult to separate the process of collaboration from the innovative results and their problem-solving effects (Torfing et al., 2020, citing Innes & Booher, 1999). That is why we use Torfing et al. (2020) criteria-based assessment tool —specifically designed to distinguish measurement of process (collaboration) from measurement of outputs (innovation)— to code policy innovation⁹. They propose four indicators to measure the degree of innovation, scored on a scale from 1 to 5:

- (1) the depth of innovation at the ideational level, understood as the degree of newness of the ideas the innovation is built on.
- (2) the depth of innovation at the level of practice, defined as the degree of newness of implementation features (mainly service production and delivery).
- (3) the character of the innovation, “defined in terms of whether the innovation as a whole can be characterised as radical or incremental” (2020:403).
- (4) the reputation of the innovation, relating to external recognition of the innovative character.

However, we only use two of them (indicators 1 and 3) since we focus on the design phase and do not consider implementation, and, thus, we couldn't gather reputation information either. The two indicators are added forming an innovation index (1-10) that is used to construct a dummy variable: policy interventions rating 6 or more are considered innovative and receive value 1, while the rest receive value 0. Examples of innovative policy interventions include the creation of a new public figure —the “industrial

⁹ Both authors coded the dependent variable separately, following Torfing et al. (2020)'s table scores included in the Appendix. They agreed on the result of the dummy variable in 93% of the cases. The resting 36 policy interventions were discussed to reach an agreement on the score.

estates agent”— that applies a “community” logic to facilitate the relationship among industrial agents and strengthen the industrial fabric of some neighbourhoods (scored 10), a regularisation programme for undocumented people (scored 9) or a “library of things”, designed to lend everyday objects (scored 8). By contrast, already-existing interventions—such as healthy eating workshops, improvements in school infrastructure or sports promotion for the elderly in public parks—have not been identified as new policy solutions. Figure 1 in the Appendix shows the distribution of the dependent variable before converting it into a dummy.

Other variables of the monitoring dataset collect different profile features of the policy interventions, some of which will be used as controls: the neighbourhood where it takes place, the main policy domain to which it relates (education, social services, urbanism, or economics) and the degree of political sensitivity (of the intervention). We would have liked to include additional controls (for example the budget of each policy intervention or some contextual variables), but the dataset did not allow for it.

Table 1. Descriptive statistics

Variable	Frequency	Number of observations
Dependent variable		
Innovative policy interventions		
<i>Not innovative</i>	74%	392
<i>Innovative</i>	26%	137
Independent variables		
H1: traditional vs. collaborative approach		
<i>Single public actor involved</i>	24%	126
<i>Two or more policy actors involved</i>	76%	403
H2: traditional vs. collaborative government		
<i>Single public actor involved</i>	35%	183
<i>Several public actors involved</i>	65%	346
H3: cross-sectoral vs. sectoral approach		
<i>Same sector</i>	40%	213
<i>Single public actor</i>	35%	184
<i>Different sectors</i>	25%	132
H4: collaborative government vs. co-creation		
<i>Public collaboration-only</i>	37%	197
<i>Single public actor</i>	23%	124
<i>Non-public actors</i>	40%	208
H5: Co-creation with each type of non-public actor		
<i>Citizens</i>		
<i>With</i>	21%	109
<i>Without</i>	79%	420
<i>Neighbourhood associations</i>		
<i>With</i>	13%	70
<i>Without</i>	87%	459
<i>Third-sector organizations</i>		
<i>With</i>	23%	119
<i>Without</i>	77%	410
<i>Companies</i>		
<i>With</i>	9%	46

<i>Without</i>	91%	483
Control variables		
Policy domain		
<i>Education</i>	24%	127
<i>Urban</i>	6%	33
<i>Social</i>	46%	243
<i>Economic</i>	24%	126
Number of management areas involved in the design process		
<i>Less than two</i>	68%	358
<i>Two or more</i>	32%	171
Neighbourhoods		
<i>Trinitat Nova</i>	10%	50
<i>Bon Pastor i Baró de Viver</i>	13%	71
<i>Besós Maresme</i>	12%	62
<i>La Teixonera i Sant Genís</i>	6%	32
<i>Trinitat Vella</i>	11%	56
<i>Raval i Gòtic Sud</i>	9%	49
<i>La Verneda i La Pau</i>	10%	55
<i>Zona Nord</i>	10%	52
<i>La Marina</i>	11%	59
<i>Roquetes</i>	8%	43

Table 1 shows the descriptive statistics of the dependent, independent and control variables. Since all of them are dummies or categorical variables, we show the categorization of policy interventions (in percentage and in number of observations) for each variable. The table shows that only 25% of policy interventions can be considered innovative. Additionally, most policy interventions are designed with more than one actor, indicating the government's commitment to collaborative approaches. Almost half of the policy interventions belong to the social sector and in almost 70% of them there are less than two management areas involved, which we use as a proxy for political sensitivity (the more areas involved, the more salient is the policy). Finally, the policies seem to be rather balanced across neighbourhoods.

4.3 Method

We use regression analysis to test the different hypotheses with the following specification¹⁰:

$$Y_{ij} = \alpha + \beta_1 x_{1ij} + \beta_2 j + \beta_k Z_n + u_i$$

where Y_{ij} is the dependent variable—a dummy variable which takes value 1 if the policy design process i in area j is defined as innovative (i.e., there is policy innovation

¹⁰ We use the linear probability model (LPM) because of two reasons. First, it provides the best linear approximation to the conditional expectation function (Angrist and Pischke, 2009) and its estimates have a very intuitive interpretation. Second, because we use neighborhood fixed effects, which may introduce bias in non-linear models (i.e., there may be an incidental parameter problem).

according to the definition in Table 1) and value 0 otherwise—and x_{ij} is the independent variable, which takes on a different shape depending on the hypothesis, representing different approaches to policy design:

- For H1, x_{1ij} is a dummy which takes value 0 if there was a single public actor involved in the policy design process i in area j , and value 1 if there were two or more policy actors involved (either public or non-public actors).
- For H2, x_{1ij} is a dummy which takes value 0 if there was a single public actor involved in the policy design process i in area j , and value 1 if there was public collaboration (collaborative or networked government), i.e., two or more *public* actors involved.
- For H3, x_{1ij} is a categorical variable which takes value 0 if there was sectoral collaboration in the policy design process i in area j (more than one public actor from the same policy domain involved), 1 if there was only one public actor involved, and 2 if there was cross-sectoral collaboration (several public actors from different policy domains involved). Thus, sectoral collaboration is the baseline category.
- For H4, x_{1ij} is a categorical variable which takes value 0 if there was *only* public collaboration in the policy design process i in area j , 1 if there was no collaboration at all (i.e., a single public actor involved) and 2 if there was public - non-public collaboration (regardless of the nature of the non-public-actor). Hence, the baseline category is 'only public collaboration'.
- For H5, the model takes this shape:

$$Y_{ij} = \alpha + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_2 j + \beta_k Z_n + u_i$$

where each variable—from x_{1ij} to x_{4ij} —is a series of dummy variables for each type of non-public actor: citizens, third sector organizations, neighbourhood associations and private companies.

We have added neighbourhood dummies j to control for neighbourhood effects and a series of controls Z , which include the sector of the policy strategy (whether it is an urban, educative, economic, or social policy) and the political relevance of the policy intervention, proxied by the number of management areas involved in the design process. In addition, for Model 5 we have added a dummy variable to control for whether one or several public actors are involved in the co-creation approach.

4.4 Results

The results show that in general, collaborative approaches to policy design are positively associated with policy innovation. Model 1 from Table 2 suggests that when policy design processes involve several actors—as opposed to traditional approaches, which involve

a single specialised administration—the probability of policy innovation increases by almost 27pp. Furthermore, Model 2 confirms that collaboration amongst public actors is enough to foster policy innovation since, comparing collaborative government to the traditional approach, the likelihood of policy innovation increases by 18pp when two or more public actors are involved.

Models 3 to 5 unpack different ways of collaborating. Focusing on public actors, collaboration amongst units from different policy sectors (the so-called cross-sectoral approach in Model 3) seems to increase the likelihood of policy innovation. Policy design processes involving actors from different sectors are 15pp more likely to result in policy innovation than policy design processes in which actors come from the same policy sector.

Finally, the inclusion of non-public actors also give rise to a higher probability of policy innovation (Models 4 and 5). Policy design processes involving non-public actors (be it citizens, private companies, third-sector organisations or neighbourhood associations) are 16pp more likely to result in policy innovation than those involving *only* public actors (Model 4). Nevertheless, when we unpack the category of non-public actors (Model 5), the results suggest that it is organized-citizenry (third-sector organisations and neighbourhood associations) which make a difference to policy innovation. Their involvement increases the probability of policy innovation by 15pp. Conversely, involving non-organized citizens does not lead to an increase in the probability of policy innovation, and the same holds for the involvement of private companies.

Table 2. Regression analysis

Dependent variable: Policy innovation (Yes =1)					
	(1)	(2)	(3)	(4)	(5)
Collaborative approach (<i>baseline: single public actor</i>)					
Several actors	0.268***				
	(0.036)				
	[0.181 –				
	0.354]				
Joined-up/Collaborative government (<i>baseline: single public actor</i>)					
Several public actors		0.166***			
		(0.039)			
		[0.085 – 0.247]			
Cross-sectoral approach (<i>baseline: same sector collab.</i>)					
Single public actor			-0.109**		
			(0.044)		
			[-0.197 – -0.020]		
Different sectors			0.149***		
			(0.052)		
			[0.055 – 0.244]		

Non-public actors' participation (<i>baseline: public collab. only</i>)					
Single public actor				-0.178***	
				(0.042)	
				[-0.275 – -0.082]	
Non-public actors				0.157***	
				(0.045)	
				[0.072 – 0.241]	
Co-creation approach (<i>dummies for each type of non-public actor</i>)					
Citizens				-0.019	
				(0.064)	
				[-0.130 - 0.092]	
Neighbourhood associations				0.151**	
				(0.068)	
				[0.037 – 0.265]	
Third-sector organisations				0.148**	
				(0.063)	
				[0.040 – 0.257]	
Companies				0.100	
				[-0.046 – 0.245]	
Controls					
Sector of the policy strategy (<i>baseline: educational sector</i>)					
<i>Urban sector</i>	0.034	0.024	0.021	0.028	0.039
	(0.078)	(0.080)	(0.077)	(0.076)	(0.081)
<i>Social sector</i>	0.083*	0.086*	0.083*	0.052	0.035
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
<i>Economic sector</i>	0.212***	0.204***	0.216***	0.207***	0.187***
	(0.054)	(0.055)	(0.055)	(0.055)	(0.056)
Number of management areas involved in the design process (<i>baseline: less than two</i>)					
+ 2	-0.041	-0.043	-0.043	-0.042	-0.052
	(0.039)	(0.040)	(0.040)	(0.039)	(0.040)
Neighbourhood dummies	Yes	Yes	Yes	Yes	Yes
Constant	-0.091	0.007	0.117	0.113	0.106
	(0.075)	(0.074)	(0.073)	(0.071)	(0.071)
Observations	529	529	529	529	529
Adjusted R ²	0.104	0.069	0.084	0.121	0.075
Note:				*p**p***p<0.01	

The results are similar when we change the dependent variable from a dummy to a continuous one¹¹, as shown in Table 2A in the Appendix. The results are robust for all models except for Model 5. In the alternative specification, policy design processes involving neighbourhood associations, and third sector organizations are not significant anymore (at a 5% significance level).

Table 3 summarises the level of support for each hypothesis.

Table 3. Summary of results

Hypothesis number	Reference category	Relevant Category explored	Coefficient and significance	Hypothesis supported yes/no
1	Single public actor	Several actors	0.268***	Yes
2	Single public actor	Several public actors	0.166***	Yes
3	Same sector collab.	Different sectors	0.149***	Yes
4	Public collab. only	Non-public collab.	0.157***	Yes
5a	Without citizens	With citizens	-0.019	No
5b	Without neighbourhood associations	With neighbourhood associations	0.151**	Yes
5c	Without Third-sector organ.	With Third-sector organ.	0.148**	Yes
5d	Without companies	With companies	0.100	No

4.5 Discussion

Enhancing collaborative innovation has become a key aspiration of public organisations in many Western countries (Sørensen & Torfing, 2016: 146, citing HM Government, 2010; Sunstein, 2012), mainly because collaboration has often been prescribed as a means for solving wicked problems and facing complexity (Scott & Thomas, 2016). Instead of simply choosing among existing tools and solutions, designers should consider the range of feasible options (Howlett, 2014) and encourage the creation of policy alternatives. And this exercise of thinking “outside the box” is supposed to be easier and more fruitful when it is a team effort. That is why this study aimed at examining this relationship—between collaboration and policy innovation—and providing quantitative evidence supporting it.

According to Sørensen and Torfing (2016), both collaborative and innovation processes are difficult to trigger in the public sector without proper management. More specifically, Crosby et al. (2016: 2) propose that, when facing wicked problems, “public managers should serve as orchestrators of networked interaction and mutual learning: acting as conveners, stewards, and catalysts of collaborative innovation”. However, for policymakers to foster “collaborative innovation”, they must first understand which actors need to get involved in the collaboration process in order to maximize the likelihood of

¹¹ Changing to a continuous variable may increase transparency of such variable; the descriptives show the distribution of the variable when it is treated as a continuous one. Nevertheless, the choice of a binomial variable for the main results seems more adequate since it allows an easier interpretation.

innovative policy outputs. By analysing and comparing the innovative potential of different collaborative arrangements we expect to provide them specific evidence to handle policy design processes and encourage policy alternatives.

However, we first reassure that collaborative approaches are indeed more likely to result in policy innovation than the traditional ones, even if collaboration takes place amongst public actors. Since policy design processes involving more than two public actors are almost 17pp more likely to result in policy innovation than those involving a single public actor, and that probability raises to 27pp when non-public actors are involved, we confirm that “innovation is seldom the result of the efforts of a single actor” (Sørensen & Torfing, 2016:152, citing Csikszentmihalyi, 1996).

Horizontal specialised sectorisation between policy domains—characteristic of traditional approaches—has been associated with policy stability (Bason, 2018). Consistently, our results confirm that **cross-sectoral approaches** to policy design are significantly linked to policy innovation. Policy design processes involving actors representing different policy sectors are 15pp more likely to result in policy innovation than those processes in which the actors involved belong to the same policy sector. This result confirms that what has been called the “silo trap” not only prevents collaboration (Sørensen & Torfing, 2016), but also innovation. This is a significant finding if we consider that both governments and public administrations continue to be organised in policy domains and that NPM reforms enhanced administrative silos (Torfing, et al., 2020).

Furthermore, the fact that the control variable “policy domain” is also significant reveals that, as Howlett (2002) has already demonstrated, policy domains are linked to specific policy subsystems, which in turn are related to policy innovation. Still and so, collaboration within a policy domain—for example between two or more education policy actors—is still more likely to result in policy innovation than traditional single-actor policy design processes. More specifically, sectoral processes are almost 11pp more likely to result in policy innovation than those involving a single public actor. That result clearly confirms that innovation is rarely generated by a single actor because it requires new ideas; and even if the societal problem continues to be framed within a specific policy sector, collaboration may foster creativity. Understanding how radical those ideas and the resulting policy alternatives might be is beyond the scope of our analysis and should be explored further.

Hence, in accordance with collaborative government defenders, **collaboration among public actors** is more than enough to foster public innovation. This is an important finding, mainly for two reasons. On the one hand, even if we have moved well past the idea that public innovation is an oxymoron, by emphasising the importance of involving non-public agents in the policymaking process, both collaborative governance and collaborative innovation researchers usually underestimate the innovative potential of public-public collaboration. Therefore, these results support the idea that “the public sector has always produced a considerable amount of policy and service innovation” (Sørensen & Torfing, 2016:149, citing Ehrenreich, 1985; Dean, 1991; Borins, 2001; Hartley, Sørensen & Torfing, 2013) by emphasising its capacity to do so on its own.

On the other hand, recognising the innovative potential of public-public collaboration implies that policy makers and policy designers committed to finding new and better solutions to societal problems do have different options when taking a collaborative approach. Frequently, literature on collaborative governance seems to claim that the more actors involved in the policy process, the better. Since the idea is to avoid privileging a specific group of actors, the solution seems to call for including all the relevant actors (Ansell & Gash, 2007; Scott & Thomas, 2016; Sørensen & Torfing, 2016). However, involving too many people may render the design process much more complex. According to Ansell et al. (2020:571), wide inclusion of actors in collaborative processes “increase transaction costs, reduce the quality of deliberation, muddy negotiations or produce “least common denominator” bargaining outcomes”. Therefore, acknowledging the innovative potential of public-public collaboration may help policy designers to choose when to open the policy design process and, especially, to whom.

In that sense, **involving non-public actors** in the policy design process increases its probability of resulting in policy innovation: design processes including non-public actors are 16pp more likely to result in policy innovation than those involving *only* two or more public agents. According to Scott and Thomas (2016:9), “by involving external stakeholders, public managers can redefine policy problems and thereby alter or expand the existing scope of action”. Non-public agents’ perspectives would help them to better understand both the problem trying to be solved and the contextual conditions, as well as to identify those solutions more likely to work on the ground and gain support for their implementation (Ansell et al., 2017; Scott & Thomas, 2016).

However, in the same way that actors involved in the policy design process have different levels of influence on policy formulation (Howlett, 2014), they seem to have different levels of capacity to innovate. That is coherent with our interest to distinguish between non-public actors and explore whose collaboration matters more in terms of innovative potential. Note that we are not discussing how to democratise the policy design process. Our main interest lies in identifying who should be part of it, in order to maximise the innovative potential of the policy output. The typology of policy actors considered—companies, third-sector organisations, neighbourhood associations and non-organised citizenship—was already predefined in the *Pla de Barris* monitoring database. Thus, we were not able to modify it.

According to the results, the involvement of **third-sector organisations and neighbourhood associations** in the policy process is what makes the difference in terms of innovative potential of the policy output. The probability of policy innovation increases by 15pp when they participate in the policy design process. By contrast, the results for companies¹² and non-organised citizens are not conclusive. This is a relevant finding since co-creation studies rarely distinguish among stakeholders or policy actors. The little existing empirical evidence supporting a positive relationship between co-creation and policy innovation (Bason, 2017; Gillinson et al., 2010; Van Gestel & Grotenbreg, 2021) does not explore each type of policy actor’s contribution.

¹² Private companies were only involved in 46 of the 529 policy design processes. Thus, the available data to account for their effect is limited.

Furthermore, available empirical evidence on co-creation is almost exclusively based on single or small-n case studies. Nevertheless, this result, and especially the role of neighbourhood associations and third sector organizations does not seem to be robust to different specifications of the dependent variable, which suggests we should take the evidence provided as a first step to encourage more research in this direction.

The robustness check does confirm the little relevance of citizens when it comes to increase policy innovation. One possible explanation for the non-significant effect of **(non-organized) citizens'** involvement is heterogeneity. Third-sector organisations tend to be experts in their field of intervention and hold well-defined values and ideas. Neighbourhood associations, on the other hand, were originally created to pressure the City Council to improve services and facilities (Martín-Gómez, 2020), and —even if they adapted to the new dynamics of policymaking— continue to hold very specific interests. Non-organised citizenry, however, is extremely diverse, not only in terms of socioeconomic profile, but also in terms of values, interests, and ideas. And collaborative policy design, as collaborative innovation, requires generating some sort of joint-ownership (Torfing, 2016): a joint commitment with shared goals, that is crucial to bridging cognitive differences. In fact, there is an academic debate on whether collaboration is associated with consensus (Straus, 2002) or constructive management of differences (Gray, 1989). We agree with the advocates of the latter approach that “total consensus is often achieved by getting everybody to agree on the lowest common denominator” (Sørensen & Torfing, 2016:155), which tends to favour incremental adjustments rather than innovation.

In sum, when we unpack the category of “non-public policy actors” we can identify that not all of them have the same innovative potential, largely due to an existing tension between innovation and agreement (Page & Kern, 2016). Some researchers have proposed to resolve this tension by adhering to “principled engagement” (Emerson et al., 2012; Emerson & Nabatchi, 2015; O’Leary & Vij, 2012; Page & Kern, 2016), understood as a process in which people with differing ideas and goals work “across their respective institutional, sectoral, or jurisdictional boundaries to solve problems, resolve conflicts or create values” (Emerson et al., 2012:10). This process, however, seems to be easily to put into practice with organized policy actors.

In this light, it is surprising that the control variable “political relevance of the policy intervention” is not significant, since, according to Ansell et al. (2017), crafting innovative policies is fundamentally a political, rather than technocratic, process. Even if the political context (debates, conflicts, pressures, short timescales...) is expected to influence the potential for policy innovation, the number of managerial departments involved in the policy design process does not appear to be linked to the innovative character of the policy output. However, it is also possible that our variable is not a good proxy for political relevance. It would have been interesting to use other variables related to political context and relevance, but unfortunately the database did not have any.

The analysis has some limitations. First, *Pla de Barris* is a municipal strategy that caters to the needs of the city of Barcelona, and therefore to a very specific context. While Barcelona might bear some resemblance to other European cities in terms of its social fabric, different varieties of capitalism and of welfare states within Europe limit the

generalisation of the results to other contexts. Second, the analysis focuses on the combination of actors participating in the design process and does not consider other aspects of collaboration that might be relevant. Third, the data may hide some information which could be interesting to unpack. We are referring to the neighbourhood fixed effects, which are statistically significant, suggesting that some factors such as historical context or demographics might be relevant to processes of policy innovation and thus could be of interest to policy makers. And fourth, the quantitative analysis does not address issues of causality. From the data available we cannot know the exact timing at which policy actors entered the policy design processes, and therefore, it could well be that policy design processes which seemed more innovative from the start attracted more policy actors. Nevertheless, we think this is quite unlikely, since when policy design processes have been in motion for a while, there are entry barriers for new actors. Additionally, it could be said that complex problems not only demand innovation but also attract more attention and thus we are observing a spurious relationship between innovation and collaboration when actually these two are brought about by the complex nature of the problem. However, a twofold argument must be considered: first, since New Public Management continues to be the default paradigm (Torfing et al., 2020), wicked problems continue to be handled traditionally in many contexts. Second, even in a context displaying New Public Governance tendencies as Barcelona city, the control variable for political relevance of the policy intervention—which should capture this “attractive” character of complex problems—was not significant.

4.6 Conclusion

Despite the increasing attention paid by both academic researchers and Western governments to collaborative innovation, public innovation remains underexplored as a tool for changing policies and services (Sørensen & Torfing, 2016). Policymakers are compelled to update their approaches to policy design to more effectively address complexity. However, before determining when and how to introduce new instruments, they must first understand the range of available approaches (Howlett, 2014).

Through our research we aim to provide specific guidance to policymakers seeking new and improved solutions to societal problems in the management of policy design processes. To do so, this article, first, contributes to confirm with empirical quantitative evidence that collaborative approaches to policy design are indeed more likely to foster policy innovation than traditional ones. As it has largely been expressed in the literature, innovation is rarely generated by a single actor because it requires new views and ideas.

Collaboration, however, can take place among different types of actors and the innovative potential of different collaborative arrangements had not been empirically analysed, nor compared. That is the main contribution of this research. As hypothesized, both cross-sectoral collaboration and collaborative government appear to be powerful approaches to foster policy innovation within the public sector. Overcoming the “silo trap” and complementing specialisation culture by undertaking deliberative processes with other policy actors seem to be an effective way to promote policy innovation. Thus, as suggested by collaborative governance researchers, those policy designers looking for

creative solutions to wicked problems can start by carefully promoting cross-boundary collaboration between either public or public and private actors in networks and partnerships (Torfing, et al., 2020). Nevertheless, in the same way that not all the actors involved in the policy design process have the same level of influence on policy formulation, not all of them have the same capacity to innovate. Therefore, those actors holding well-defined values, interests and ideas should be prioritised when the innovative potential wants to be maximised.

Finally, we think this study points to avenues for further research, in at least three directions. First, innovation and quality are far from being synonyms. A policy can be innovative and at the same time be poorly designed. Hence, considering that bad designs are one of the main causes of policy failure in the implementation phase (Ansell et al., 2017), more attention should be paid to the relationship between innovation and quality at the time of designing the policy. Second, innovation does not guarantee avoiding policy failure. We have focused on the phase of policy design, and therefore, more research on the relationship between innovation and policy success would complement the evidence provided in this article to guide policymakers and especially policy designers in the management of complexity. And third, having proved the relevance of non-public actors, we think it could be of interest to explore the innovative potential of each type of actor in order to better understand their influence and thus make more precise recommendations to policy designers.

4.7 References

- Agranoff, Robert. (2003). *Leveraging Networks: A Guide for Public Managers Working across Organizations*. New Ways to Manage Series. IBM Endowment for The Business of Government. <http://www.businessofgovernment.org/report/leveraging-networks-guide-public-managers-working-across-organizations>
- Angrist, J., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (1st edition). Princeton University Press.
- Ansell, C., Doberstein, C., Henderson, H., Siddiki, S., & Hart, P. 't. (2020). Understanding inclusion in collaborative governance: A mixed methods approach. *Policy and Society*, 39(4), 570–591. <https://doi.org/10.1080/14494035.2020.1785726>
- Ansell, C., & Gash, A. (2007). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory*, 18(4), 543–571. <https://doi.org/10.1093/jopart/mum032>
- Ansell, C., Sørensen, E., & Torfing, J. (2017). Improving policy implementation through collaborative policymaking. *Policy and Politics*, 45(3), 467–486. <https://doi.org/10.1332/030557317X14972799760260>
- Ansell, C., & Torfing, J. (2014). *Public Innovation through Collaboration and Design* (p. 246). <https://doi.org/10.4324/9780203795958>
- Ansell, C., & Torfing, J. (2021). Co-creation: The new kid on the block in public governance. *Policy & Politics*, 49(2), 211–230. <https://doi.org/10.1332/030557321X16115951196045>

- Arbolino, R., Boffardi, R., & De Simone, L. (2019). Which are the Factors Influencing Innovation Performances? Evidence from Italian Cohesion Policy. *Social Indicators Research*, 146(1–2), 221–247. <https://doi.org/10.1007/s11205-018-1904-5>
- Arundel, A., Bloch, C., & Ferguson, B. (2019). Advancing innovation in the public sector: Aligning innovation measurement with policy goals. *Research Policy*, 48(3), 789–798. <https://doi.org/10.1016/j.respol.2018.12.001>
- Bason, C. (2010). *Leading Public Sector Innovation: Co-creating for a Better Society*. Policy Press.
- Bason, C. (2017). *Leading Public Design. Discovering Human-Centred Governance* (Policy Press. University of Bristol). <https://policy.bristoluniversitypress.co.uk/leading-public-design>
- Bason, C. (2018). *Leading Public Sector Innovation 2E: Co-creating for a Better Society*. Policy Press.
- Bentzen, T. Ø., Sørensen, E., & Torfing, J. (2020). Strengthening public service production, administrative problem solving, and political leadership through co-creation of innovative public value outcomes? *The Innovation Journal: The Public Sector Innovation Journal*, 25(1).
- Bhatta, G. (2003). *Revisiting the Issue of Risks in Innovation in the Public Sector*. 8, 12.
- Bingham, L. B., Nabatchi, T., & O’Leary, R. (2005). The New Governance: Practices and Processes for Stakeholder and Citizen Participation in the Work of Government. *Public Administration Review*, 65(5), 547–558.
- Bommert, B. (2010). Collaborative Innovation in the Public Sector. *International Public Management Review*, 11(1), 1.
- Borins, S. (1998). *Innovating with Integrity*. Georgetown University Press. <https://www.perlego.com/book/949540/innovating-with-integrity-how-local-heroes-are-transforming-american-government-pdf>
- Borins, S. (2001). Encouraging Innovation in the Public Sector. *Journal of Intellectual Capital*, 2, 310–319. <https://doi.org/10.1108/14691930110400128>
- Bourgon, J. (2011). *A New Synthesis of Public Administration: Serving in the 21st Century*. School of Policy Studies, Queen’s University.
- Brandsen, T., Steen, T., & Verschuere, B. (2018). Co-Creation and Co-Production in Public Services: Urgent Issues in Practice and Research. In *Co-Production and Co-Creation*. Routledge.
- Crosby, B. C., & Bryson, J. M. (2010). Integrative leadership and the creation and maintenance of cross-sector collaborations. *The Leadership Quarterly*, 21(2), 211–230. <https://doi.org/10.1016/j.leaqua.2010.01.003>
- Crosby, B., Hart, P., & Torfing, J. (2016a). Public value creation through collaborative innovation. *Public Management Review*, 19, 1–15. <https://doi.org/10.1080/14719037.2016.1192165>

Crosby, B., Hart, P., & Torfing, J. (2016b). Public value creation through collaborative innovation. *Public Management Review*, 19, 1–15. <https://doi.org/10.1080/14719037.2016.1192165>

Daglio, M., Gerson, D., & Kitchen, H. (2014). *Building Organisational Capacity for Public Sector Innovation* (Background Paper Prepared for the OECD Conference “Innovating the Public Sector: From Ideas to Impact”). <https://www.oecd.org/innovating-the-public-sector/Background-report.pdf>

Damanpour, F. (1991a). Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators. *The Academy of Management Journal*, 34(3), 555–590. <https://doi.org/10.2307/256406>

Damanpour, F. (1991b). Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators. *The Academy of Management Journal*, 34(3), 555–590. <https://doi.org/10.2307/256406>

Dente, B., Bobbio, L., & Spada, A. (2012). Government or Governance of Urban Innovation? *DisP - The Planning Review*, 41, 41–52. <https://doi.org/10.1080/02513625.2005.10556931>

Doz, Y., & Koskonen, M. (2014). *Governments for the future: Building the strategic and agile state*. Sitra. <https://media.sitra.fi/2017/02/23222725/Selvityksia80.pdf>

Eggers, W., & O’Leary, J. (2009). *If We Can Put a Man on the Moon... Getting Big Things Done in Government*. Harvard Business Press. <https://www.scribd.com/book/353493797/If-We-Can-Put-a-Man-on-the-Moon-Getting-Big-Things-Done-in-Government>

Eggers, W., & Singh, S. (2009). *The Public Innovator’s Playbook: Nurturing Bold Ideas in Government* (Harvard Kennedy School’s Ash Center). <https://ash.harvard.edu/links/public-innovators-playbook-nurturing-bold-ideas-government>

Emerson, K., & Nabatchi, T. (2015). *Collaborative Governance Regimes*. Georgetown University Press. <https://www.jstor.org/stable/j.ctt19dzcvf>

Emerson, K., Nabatchi, T., & Balogh, S. (2012). An Integrative Framework for Collaborative Governance. *Journal of Public Administration Research and Theory*, 22(1), 1–29. <https://doi.org/10.1093/jopart/mur011>

Fung, A., & Wright, E. O. (2003). Thinking about Empowered Participatory Governance. In *Deepening Democracy*. Verso. /paper/Thinking-about-Empowered-Participatory-Governance-*-Fung-Wright/0427afba65347927bfd48922ec083abb9859eaa8

Geyer, R., & Rihani, S. (2010). *Complexity and Public Policy: A New Approach to 21st Century Politics, Policy And Society*. Routledge. <https://www.routledge.com/Complexity-and-Public-Policy-A-New-Approach-to-21st-Century-Politics-Policy/Geyer-Rihani/p/book/9780415556637>

Gillinson, S., Horne, M., & Baeck, P. (2010). *Radical Efficiency: Different, Better, Lower Cost Public Services*. Innovation Unit, The Lab, and Nesta. <chrome-extension://efaidnbmninnibpcapcglclefindmkaj/http://www.networkforeurope.eu/files/files/radical-efficiency-nesta-180610.pdf>

Gray, B. (1989). *Collaborating: Finding Common Ground for Multiparty Problems*. Wiley.

Hartley, J. (2005). Innovation in Governance and Public Services: Past and Present. *Public Money & Management*, 25(1), 27–34. <https://doi.org/10.1111/j.1467-9302.2005.00447.x>

Hartley, J., Sørensen, E., & Torfing, J. (2013). Collaborative Innovation: A Viable Alternative to Market Competition and Organizational Entrepreneurship. *Public Administration Review*, 73(6), 821–830. <https://doi.org/10.1111/puar.12136>

Head, B. W. (2022). *Wicked Problems in Public Policy: Understanding and Responding to Complex Challenges*. Palgrave Macmillan.

Howlett, M. (2002). Do Networks Matter? Linking Policy Network Structure to Policy Outcomes: Evidence from Four Canadian Policy Sectors 1990-2000. *Canadian Journal of Political Science/Revue Canadienne de Science Politique*, 35(2), 235–267. <https://doi.org/10.1017/S0008423902778232>

Howlett, M. (2014). From the ‘old’ to the ‘new’ policy design: Design thinking beyond markets and collaborative governance. *Policy Sciences*, 47(3), 187–207. <https://doi.org/10.1007/s11077-014-9199-0>

Innes, J. E., & Booher, D. E. (1999). Consensus Building and Complex Adaptive Systems. *Journal of the American Planning Association*, 65(4), 412–423. <https://doi.org/10.1080/01944369908976071>

Kanter, R. (1988). When a Thousand Flowers Bloom: Structural, Collective, and Social Conditions for Innovation in Organization. *Research in Organizational Behaviour*, 10, 196–211. <https://doi.org/10.1016/B978-0-7506-9749-1.50010-7>

Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Viking.

Kimberly, J. R., & Evanisko, M. J. (1981). Organizational Innovation: The Influence of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations. *The Academy of Management Journal*, 24(4), 689–713. <https://doi.org/10.2307/256170>

Krogh, A., & Torfing, J. (2015). Leading collaborative innovation. In A. Agger, A. H. Damgaard, A. Krogh, & E. Sørensen (Eds.), *Collaborative Governance and Public Innovation in Northern Europe* (pp. 91–110). Bentham Science Publishers.

Martín-Gómez, A. (2020). Participación ciudadana deliberativa en el urbanismo: «Repensemos el 22@» en Barcelona. *Gestión y Análisis de Políticas Públicas*, 24, 120–138.

Mulgan, G. (2007). Ready or not? Taking innovation in the public sector seriously. *NESTA, Provocation*, 03, 40.

Mulgan, G. (2009). *The Art of Public Strategy: Mobilizing Power and Knowledge for the Common Good*. OUP Oxford.

Mulgan, G., & Albury, D. (2003). *Innovation in the public sector* (1.9; Working Paper). UK Cabinet Office. <https://www.alnap.org/system/files/content/resource/files/main/innovation-in-the-public-sector.pdf>

Nabatchi, T., Sancino, A., & Sicilia, M. (2017). Varieties of Participation in Public Services: The Who, When, and What of Coproduction. *Public Administration Review*. <https://doi.org/10.1111/puar.12765>

Nambisan, S. (2008). *Transforming Government Through Collaborative Innovation* (p. 44). IBM Center for the Business of Government.

Newman, J., Raine, J., & Skelcher, C. (2001). Developments: Transforming Local Government: Innovation and Modernization. *Public Money & Management*, 21(2), 61–68. <https://doi.org/10.1111/1467-9302.00262>

O'Leary, R., & Vij, N. (2012). Collaborative Public Management Where Have We Been and Where Are We Going? *The American Review of Public Administration*, 42, 507–522. <https://doi.org/10.1177/0275074012445780>

Osborne, S. P. O. (2006). The New Public Governance? *Public Management Review*, 8(3), 377–387. <https://doi.org/10.1080/14719030600853022>

Page, S., & Kern, M. (2016). *Innovation and Agreement? Tensions among the Outputs of Collaborative Governance* (SSRN Scholarly Paper ID 2767820). Social Science Research Network. <https://papers.ssrn.com/abstract=2767820>

Paquet, G. (2009). *Scheming virtuously: The road to collaborative governance*. Invenire Books.

Parés, M., Castellà, C., & Brugada, M. (2015). *Repensant la participació de la ciutadania al món local*.

Paz, B., & Fontaine, G. (2018). A Causal Mechanism of Policy Innovation: The Reform of Colombia's Oil-Rents Management System. *Revista de Estudios Sociales*, 63, 2–19. <https://doi.org/10.7440/res63.2018.01>

Rhodes, R. (2007). Understanding Governance: Ten Years On. *Organization Studies*, 28(8), 1243–1264. <https://doi.org/10.1177/0170840607076586>

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169. <https://doi.org/10.1007/BF01405730>

Roberts, N., & Bradley, R. (1991). Stakeholder Collaboration and Innovation: A Study of Public Policy Initiation at the State Level. *The Journal of Applied Behavioral Science*, 27, 209–227. <https://doi.org/10.1177/0021886391272004>

Roberts, N., & King, P. (1996). Transforming Public Policy: Dynamics of Policy Entrepreneurship and Innovation. *Undefined*. <https://www.semanticscholar.org/paper/Transforming-Public-Policy%3A-Dynamics-of-Policy-and-Roberts-King/8399f6b1701a98588476e73b8c89f97b388310f6>

Scott, T., & Thomas, C. (2016). Unpacking the Collaborative Toolbox: Why and When Do Public Managers Choose Collaborative Governance Strategies?: Unpacking the Collaborative Toolbox. *Policy Studies Journal*, 45. <https://doi.org/10.1111/psj.12162>

Sørensen, E., & Torfing, J. (2011). Enhancing Collaborative Innovation in the Public Sector. *Administration & Society*, 43(8), 842–868. <https://doi.org/10.1177/0095399711418768>

Sørensen, E., & Torfing, J. (2016). *Enhancing Public Innovation through Collaboration, Leadership and New Public Governance* (pp. 145–169). https://doi.org/10.1007/978-1-137-50680-1_8

Straus, D. (2002). *How to Make Collaboration Work: Powerful Ways to Build Consensus, Solve Problems, and Make Decisions*. Berrett-Koehler.

Torfing, J. (2016). *Collaborative Innovation in the Public Sector*. Georgetown University Press. <https://www.jstor.org/stable/j.ctt1h64pfm>

Torfing, J., Andersen, L. B., Greve, C., & Klausen, K. K. (2020). New Public Governance. In *Public Governance Paradigms* (pp. 125–150). Edward Elgar Publishing. http://www.elgaronline.com/display/9781788971218/13_chapter8.xhtml

Torfing, J., Ferlie, E., Jukić, T., & Ongaro, E. (2021). A theoretical framework for studying the co-creation of innovative solutions and public value. *Policy & Politics*, 49(2), 189–209. <https://doi.org/10.1332/030557321X16108172803520>

Torfing, J., Krogh, A. H., & Ejrnæs, A. (2020). Measuring and assessing the effects of collaborative innovation in crime prevention. *Policy & Politics*, 48(3), 397–423. <https://doi.org/10.1332/030557320X15788414270675>

Van de Ven, A. H., Polley, D., Garud, R., & Venkataraman, S. (2008). *The Innovation Journey*. Oxford University Press.

Van Gestel, N., & Grotenbreg, S. (2021). Collaborative governance and innovation in public services settings. *Policy & Politics*, 49, 249–265. <https://doi.org/10.1332/030557321X16123785900606>

Verschuere, B., Brandsen, T., & Pestoff, V. (2012). Co-production: The State of the Art in Research and the Future Agenda. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 23(4), 1083–1101.

Wagenaar, H. (2007). Governance, Complexity, and Democratic Participation: How Citizens and Public Officials Harness the Complexities of Neighborhood Decline. *The American Review of Public Administration*, 37(1), 17–50. <https://doi.org/10.1177/0275074006296208>

Weber, E. P., & Khademian, A. M. (2008). Wicked Problems, Knowledge Challenges, and Collaborative Capacity Builders in Network Settings. *Public Administration Review*, 68(2), 334–349. <https://doi.org/10.1111/j.1540-6210.2007.00866.x>

4.8 Appendix

Appendix 1: Score tables measuring innovation (Torfing et al., 2020)

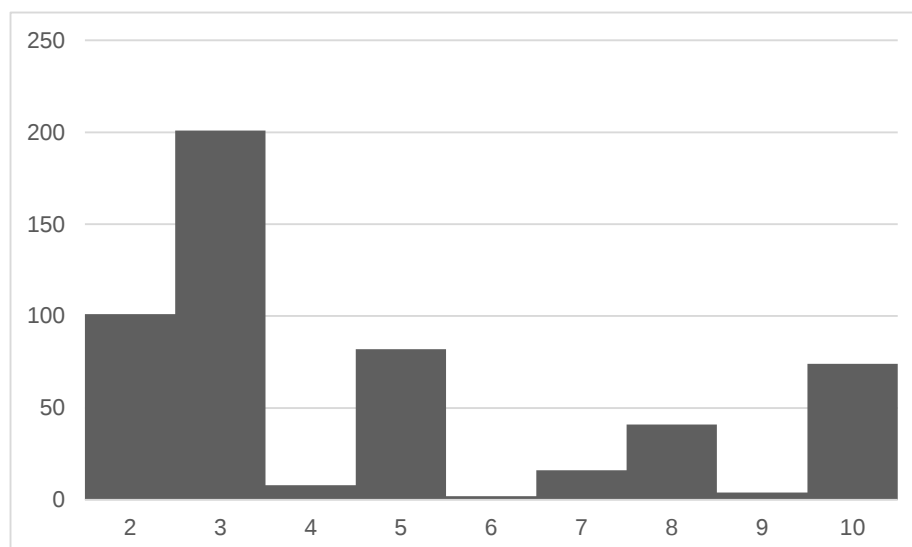
Score Table 1: The depth of innovation at the ideational level

Points	Items/criteria
5	The innovation builds on a completely new and different programme or change theory that does not merely propose new targets and the means to achieve them, but also changes the underlying understanding of the problems and challenges faced.
4	The innovation builds on a modified programme or change theory that both proposes new targets and the means to achieve them.
3	The innovation introduces new ideas and forms of knowledge which enable existing targets to be achieved using completely new means.
2	The innovation builds on new ideas that originate either internally or externally, and which change the form and content of existing solutions.
1	The innovation builds mainly on old ideas that are combined with new and different means which create new results and effects.

Score Table 2: The character of the innovation

Points	Items/criteria
5	The innovation constitutes a radical break with established practice in the area in question, as well as with underlying assumptions and actors' role perceptions; and it requires an extensive and complex change process.
4	The innovation builds on new ideas and significantly changes established practices and assumptions in the field.
3	The innovation challenges some established practices and assumptions in the field while leaving others unaffected.
2	The innovation constitutes a clear break with existing ways of doing things, but builds mainly on existing elements.
1	The innovation is small-scale and only differs to a minor extent from ongoing everyday improvements aimed at improving the area in question.

Appendix 2: Distribution of the dependent variable before converting it into a dummy



Appendix 3: Regression analysis – innovation as a continuous variable (Table 2A)

Dependent variable: Policy innovation (continuous from 1 to 10)					
	(1)	(2)	(3)	(4)	(5)
Collaborative approach (<i>baseline: single public actor</i>)					
Several actors	1.355***				
	(0.238)				
Joined-up/Collaborative government (<i>baseline: single public actor</i>)					
Several public actors		0.846***			
		(0.248)			
Cross-sectoral approach (<i>baseline: same sector collab.</i>)					
Single public actor			-0.464*		
			(0.273)		
Different sectors			0.996***		
			(0.315)		
Non-public actors' participation (<i>baseline: public collab. only</i>)					
Single public actor				-0.884***	
				(0.279)	
Non-public actors				0.847***	
				(0.279)	

Co-creation approach (<i>dummies for each type of non-public actor</i>)					
Citizens	-0.044				
	(0.387)				
Neighbourhood associations	0.659				
	(0.411)				
Third-sector organisations	0.733*				
	(0.376)				
Companies	0.583				
	(0.520)				
Controls					
Sector of the policy strategy (<i>baseline: educational sector</i>)					
<i>Urban sector</i>	0.532	0.477	0.460	0.499	0.565
	(0.437)	(0.447)	(0.428)	(0.432)	(0.452)
<i>Social sector</i>	0.619**	0.636**	0.612**	0.450*	0.384
	(0.256)	(0.256)	(0.256)	(0.254)	(0.256)
<i>Economic sector</i>	1.821***	1.783***	1.863***	1.795***	1.705***
	(0.345)	(0.348)	(0.342)	(0.346)	(0.353)
Number of management areas involved in the design process (<i>baseline: less than two</i>)					
<i>More than 2 management areas</i>	-0.176	-0.185	-0.185	-0.180	-0.234
	(0.249)	(0.253)	(0.250)	(0.247)	(0.251)
Neighbourhood dummies	Yes	Yes	Yes	Yes	Yes
Constant	2.500	2.988***	3.460***	3.509***	3.510***
	(0.458)	(0.443)	(0.426)	(0.437)	(0.424)
Observations	529	529	529	529	529
Adjusted R ²	0.100	0.078	0.096	0.114	0.079
<i>Note:</i>				*p**p***p<0.01	

5. Conclusiones

5.1 Principales resultados

El objetivo de esta tesis era explorar la contribución tanto de la evaluación como del diseño colaborativo a la innovación de políticas públicas, y con ello profundizar en el papel que juegan y pueden jugar en el abordaje de problemas sociales complejos. La intención era explorar su potencial por separado, pero también poner ambas estrategias en relación, no tanto en perspectiva comparada sino más bien con la intención de identificar posibles sinergias y/o complementariedades.

El **papel que el diseño colaborativo puede jugar a la hora de innovar en términos de política pública** se ha abordado de forma directa en el Capítulo 4 (tercer artículo). Los resultados muestran cómo, de forma consistente, todos los modelos de colaboración explorados —independientemente de los actores involucrados— obtienen mejores resultados en términos de innovación que el diseño a menos de una única administración pública. Así, la investigación aporta evidencia empírica en favor de apostar por fórmulas colaborativas a la hora de diseñar políticas públicas, como estrategia para obtener nuevas —y ojalá mejores— soluciones a los problemas complejos. Adicionalmente, Head (2022) apunta que los enfoques colaborativos también resultan útiles para mejorar la capacidad de desarrollar políticas públicas (*policy development capacity*) en condiciones de riesgo e incertidumbre (Head, 2022). En ese sentido, Ansell y Gash (2018) argumentan que las “plataformas” colaborativas —entendidas como espacios de diálogo para una discusión constante sobre el diseño y la implementación de las políticas públicas— permiten identificar procesos pragmáticos y adaptativos que facilitan el abordaje de necesidades complejas en condiciones de cambio constante.

La **contribución de la evaluación a la innovación de políticas** es menos directa y más controvertida, pues en ocasiones —según el enfoque desde el que se trabaja— parece asumir una influencia directa de la evidencia en el *policymaking*. Quienes —dentro y fuera de las administraciones públicas— se dedican a promover la cultura de la evaluación se encuentran con resistencias, especialmente en contextos con menor tradición de evaluación como Cataluña. Lo cual en muchos casos implica que deben adoptar un rol pedagógico, que los lleva a explicar constantemente la importancia, la pertinencia y el potencial transformador de la política informada en evidencia. Sin embargo, esta necesidad permanente de convencer —para superar las resistencias derivadas de la falta de cultura de evaluación— puede llevar a una cierta simplificación del papel de la evidencia en el funcionamiento real de la política pública y su capacidad de contribuir a la innovación. Una simplificación que se asocia con un enfoque racionalista del *policymaking* y que ha sido bastante cuestionada a nivel académico.

Los principales argumentos se pueden sintetizar en dos grandes ideas. La primera es que **la evidencia no es la única forma de conocimiento que tienen en cuenta los policymakers** (Sanderson, 2009), de manera que —en el mejor de los casos— su influencia siempre se va a dar en diálogo con otras fuentes de información, como la experiencia, la inercia o los hábitos, los valores, las intuiciones, el mandato político, etc. Además, según Cairney (2019) **ni siquiera es la forma de conocimiento que**

priorizan. Según el autor (2019, referenciando a Cairney and Kwiatkowski 2017; Kahneman 2012; Haidt 2001; Alter and Oppenheimer 2009), cuando los *policymakers* se ven incapaces de tomar en consideración toda la información relevante a su alcance utilizan dos “atajos”: (1) formas “racionales” para definir cuál es la mejor evidencia y las mejores fuentes de evidencia, y (2) formas “irracionales” de entender los problemas de política pública, basándose en emociones, hábitos y especialmente en creencias. Y a menudo la acumulación de conocimiento científico (o evidencia rigurosa) resulta más bien abrumadora —“demasiados problemas a los que prestar atención, demasiadas soluciones por considerar y demasiadas decisiones que tomar” de forma que las formas “irracionales” acostumbran a imponerse (Cairney, 2019:4). En ese sentido, Newman y Head (2017) apuntan que el hecho de que “todos los problemas puedan mostrar tendencias “malditas”, independientemente de la cantidad de evidencia científica disponible o en manos de los *decision-makers*, muestra que la acumulación de información (o de evidencia) nunca será, por sí misma, suficiente para resolver problemas malditos. Es más, según Cairney y Oliver (2017), a menudo la demanda (de los *policymakers*) de evidencia se produce cuando ya tienen un problema concreto pre-definido (en base a fuentes “irracionales”: creencias, inercias, valores...) y necesitan buscar soluciones.

La segunda idea es que asumir una relación directa entre la evidencia y la toma de decisión —lo que sería un uso instrumental cuando la evidencia ha sido producida mediante una evaluación— supone una **comprensión linear y seguramente simplificada del ciclo de funcionamiento de la política pública** (el llamado *policy cycle*) (Cairney, 2019). Linear porque asume que las etapas o fases del ciclo se producen de forma secuencial, en vez de reconocer —como apuntan buena parte de la literatura tanto teórica como empírica— que se superponen e influyen mutuamente (Jann & Wegrich, 2007). Y simplificada en el sentido de que no permite reconocer la multiplicidad de actores que —tal como las propuestas de la Nueva Gobernanza Pública, la gobernanza colaborativa o la gobernanza en red ponen de manifiesto— influyen y participan (de manera directa o indirecta) en el *policymaking*. De hecho, los investigadores e investigadoras son solamente uno más de esos múltiples actores que inciden o buscan incidir en el *policymaking* (Cairney & Oliver, 2017). Y a veces ni siquiera está claro quién son los *policymakers*, las personas que toman las decisiones relacionadas con el diseño y la implementación de una política pública concreta (de Marchi et al., 2016).

Olejniczak et al. (2020) añaden un tercer argumento. Quienes diseñan las políticas públicas (los llamados *policy designers*) tienden a **asumir un efecto directo o automático de las intervenciones sobre los resultados**: la llamada “caja negra” de la política pública (Pawson, 2013). Es decir, en el proceso de diseño suelen ignorar la existencia de una serie de mecanismos —y de factores contextuales— que posibilitan que se produzca ese resultado. Esta comprensión limitada del funcionamiento de las intervenciones condiciona, a su vez, el potencial de la evidencia —asumiendo que sea tenida en cuenta en el diseño de esas intervenciones.

Todo ello apunta que “el *policymaking* es un proceso altamente político y politizado, lo cual comporta que el conocimiento experto puede ser fácilmente ignorado” (Head, 2022:17). Eso no significa que la idea de análisis informados en evidencia no sea pertinente ni relevante. De hecho, los resultados del Capítulo 3 (artículo 2) muestran

que cuando se dan determinadas condiciones la evidencia puede lograr incidir en el *policy process*. Sin embargo, **no se puede desligar el *policymaking* del sistema de gobernanza, como si se tratara de un proceso aislado y tecnocrático**, que puede ser comprendido, analizado y gestionado con independencia de la política. Parsons (2002:2) lleva esta crítica al extremo al afirmar lo siguiente:

“EBPM must be understood as a project focused on enhancing the techniques of managing and controlling the policy making process as opposed to either improving the capacities of the social sciences to influence the ‘practices of democracy’ as envisaged by Lasswell, or facilitating the kind of systems thinking advocated by Schön in the 1970s and by more recent students of ‘complexity’”.

Mulgan (2005:224) se suma a la crítica al afirmar que “en una democracia, la gente, así como los políticos que les representan, tienen todo el derecho a ignorar la evidencia”, y de esta manera separa completamente evidencia o rigor científico y democracia. Sin embargo, más que apuntar y reconocer los límites de la evidencia, el objetivo —por lo menos en el marco de esta tesis— es adquirir una mejor comprensión —más profunda, pero sobre todo más realista— de su relación con el *policymaking*. Y en ese sentido Sanderson (2009:699) mantiene un discurso más constructivo y conciliador, que resume muy bien con la siguiente frase: “*policy making is more a ‘craft’ than a science; the ‘art of the possible’ rather than the ‘art of the optimum’*”.

Más concretamente, el autor (2009:700) apunta que es necesario “revisar la lógica y los supuestos fundacionales del movimiento de políticas basadas en evidencia para determinar si existe un modelo “normativo” mejor para guiar el *policymaking* en este mundo cada vez más complejo y ambiguo, que los gobiernos tienen la responsabilidad de mejorar”. El modelo que propone se basa en combinar el conocimiento existente sobre sistemas complejos adaptativos (*complex adaptive systems*) y la tradición filosófica del pragmatismo —especialmente las ideas de John Dewey— y recibe el nombre de “*policymaking* inteligente” (como alternativa a *policymaking* basado en evidencia). Sanderson (2009) apuesta por **poner en el centro el aprendizaje** como forma de encajar las políticas públicas en contextos complejos. Propone entenderlas como hipótesis que hay que testear en la práctica, a partir de evaluaciones rigurosas, para aprender de estos procesos y generar una inteligencia colectiva que pueda servir para pensar las futuras políticas y decisiones. Testear en el sentido dinámico de evaluar rápidamente el funcionamiento de las intervenciones para ajustar procesos y adaptarse a los cambios constantes e impredecibles de un contexto cambiante. Colectiva porque el aprendizaje sobre políticas públicas (*policy learning*) requiere —según el autor— fortalecer los procesos y canales de comunicación y diálogo dentro del gobierno (de forma intersectorial) y más allá de este, con la sociedad civil.

De esta forma, la propuesta de Sanderson (2009) **conjuga el papel de la evidencia, y la evaluación —rigurosa e institucionalizada—, con el de los enfoques colaborativos**, inicialmente explorados por separado en esta tesis. Un posicionamiento que encaja perfectamente con la voluntad de esta investigación de buscar y concretar sinergias entre ambas, y que también sugiere Head (2022:126), cuando afirma que “la búsqueda de nuevas soluciones debe avanzar de manera que combine buenos procesos (inclusivos), buena evidencia (una amplia gama de conocimientos y experiencias) y un ajuste iterativo en la implementación de los programas

(adaptabilidad)". Y es que al final, si algo hemos aprendido durante estos 50 años de "problemas malditos", es que ninguna estrategia será suficiente por sí misma.

El encaje podría encontrarse en la idea de "aprendizaje", que se encuentra en la base tanto de la evidencia como de la colaboración. Cuando la evaluación sirve para generar un aprendizaje social de carácter reflexivo —como propone (Sanderson, 2002)—, y por lo tanto se plantea en el marco de un dialogo constante en el interior del gobierno, y entre éste y la sociedad civil, permite tender un puente entre la complejidad social y los requerimientos de una política pública efectiva.

5.2 Relevancia de la tesis

Esta tesis contribuye a llenar tres vacíos de conocimiento específicos —expuestos en cada uno de los artículos / capítulos— y por lo tanto relevantes en sí mismos. Los tres artículos enriquecen la literatura existente en diferentes ámbitos de conocimiento (la institucionalización de la evaluación, el uso de las evaluaciones y la innovación colaborativa en políticas públicas). Sin embargo, tal como se anticipaba en la Introducción, la relevancia de cada una de estas contribuciones reside en aspectos distintos.

La contribución del **Capítulo 4** es, sobre todo, metodológica, pues el hecho de que no existiera **evidencia empírica que comparara el potencial de innovación de diferentes tipos de enfoques colaborativos** también es especialmente relevante. Si bien existía evidencia empírica para sustentar la relación entre colaboración e innovación de políticas públicas esta se fundamentaba eminentemente en estudios de caso y estudios comparados de tipo cualitativo. Por lo tanto, **complementar esta evidencia con un análisis cuantitativo** supone una aportación relevante para una literatura que está creciendo mucho y muy rápido en términos de desarrollo teórico pero que suele encontrarse con dificultades para trabajar a partir de muestras grandes como la utilizada.

Por lo que respecta a la evaluación de políticas públicas y su relación con la innovación, la contribución de esta tesis tiene un carácter distinto. Dentro de la literatura sobre evaluación de políticas existen determinados temas, como por ejemplo el uso de las evaluaciones, que han sido ampliamente desarrollados a nivel teórico y otros —como es el caso de la institucionalización de la evaluación— que son todavía relativamente nuevos y por lo tanto están en proceso de asentar unas bases conceptuales sólidas. En cualquier caso, existe una distancia importante entre quienes conceptualizan e investigan sobre el papel de la evaluación y el personal de las administraciones públicas que —como parte de sus funciones— encarga evaluaciones (los *policymakers*) y que luego decide qué hacer con esa información (*policydecisors*). **Esta tesis contribuye a acercar esos dos públicos** —que son además los dos principales agentes que generan conocimiento (de distintos tipos) sobre evaluación— en la medida en que **utiliza la investigación científica para proporcionar orientaciones específicas a los *policymakers* y *policydecisors*, lo cual proporciona a la investigación un carácter aplicado.**

Del primer artículo (**Capítulo 2**) se desprenden orientaciones concretas que pueden tener en cuenta quienes se encargan de construir los sistemas de evaluación y por lo tanto de impulsar la institucionalización de la evaluación en las administraciones públicas. También se deriva una advertencia relevante: si las normas, estructuras, políticas y buenas prácticas —los principales mecanismos de institucionalización— no se acompañan de valores, la evaluación corre el riesgo de burocratizarse.

En ese sentido es importante destacar que, en Cataluña —contexto en el que se desarrolla esta tesis— una parte muy importante de estas funciones se asumen desde Ivàlua y desde el Área de Evaluación Económica de la Dirección General de Presupuestos del Departamento de Economía y Finanzas, de la que depende orgánicamente Ivàlua. Así, que esta tesis se realice “desde Ivàlua” (por la modalidad de Doctorado Industrial) aumenta sustancialmente el potencial transformador del conocimiento producido, pues facilita que pueda ser directa y debidamente transferido a quienes deben tomar estas decisiones.

La situación respecto al segundo artículo (**Capítulo 3**) es similar. La investigación identifica condiciones necesarias y suficientes para el uso instrumental de las evaluaciones; condiciones cuya implementación depende (1) de quienes encargan evaluaciones y (2) de quienes las ejecutan. Si bien Ivàlua no es la única institución que realiza evaluaciones en Cataluña, sí es —por Acuerdo de Gobierno— el proveedor principal de evaluación para los Departamentos de la Generalitat, lo cual significa que realiza una parte muy importante de las evaluaciones estratégicas. Al mismo tiempo, su singularidad en términos funcionales y organizacionales (siendo un organismo independiente adscrito a la Dirección General de Presupuestos) le otorga importantes ventajas y capacidades a la hora de negociar con los Departamentos las condiciones en las que se realizan esas evaluaciones. Así pues, que esta investigación se realice “desde Ivàlua” garantiza que sus resultados sean tenidos en cuenta, lo cual, a su vez, aumenta las posibilidades de que los resultados de las evaluaciones sean tenidos en cuenta.

Más allá del ámbito de la evaluación, la contribución del tercer artículo (**Capítulo 4**) en términos de recomendaciones para los *policymakers* también es sustantiva y específica. Quienes quieren encontrar soluciones de política pública innovadoras y confían en los enfoques colaborativos para hacerlo tienen diversas opciones: o bien trabajar de forma intersectorial y por lo tanto interdepartamental, o bien —si quieren maximizar el potencial de innovación— invitar a otros agentes a participar en diseño de la solución. Esos “otros agentes” no públicos pueden ser empresas, organizaciones del tercer sector, universidades, asociaciones vecinales o incluso ciudadanía no organizada; sin embargo, si no quieren o no se dan las condiciones para desplegar un proceso de co-diseño plural, y quieren maximizar el potencial innovador, deberían priorizar a aquellos agentes que tienen intereses e ideas cohesionados y bien definidos (sobre el problema a abordar). Esto hace especialmente relevante la participación de las organizaciones del tercer sector.

Si bien en este caso la posibilidad de acercarles o transferirles directamente este conocimiento es menor, pues el diseño de políticas públicas es una competencia mucho más dispersa en términos organizacionales y no se encuentra entre las funciones de

Ivàlua, los resultados son suficientemente específicos para que puedan ayudar a informar la toma de decisiones de los *policy designers*.

Adicionalmente, los resultados de esta tesis contribuyen a evidenciar que, tal como apunta Head (2022:129) “la búsqueda de la innovación en políticas públicas exige claramente el **desarrollo de nuevas habilidades y capacidades**” **por parte de los *policymakers***. Habilidades relacionadas tanto con la evaluación y la consideración de la evidencia científica en la toma de decisiones (análisis de datos, análisis prospectivo, planificación de escenarios o diseño experimental, según apunta Head) como con “nuevos métodos de facilitación, para generar nuevas ideas y promover el pensamiento creativo a través de procesos multi-actor”. En este sentido los resultados presentados contribuyen a visibilizar la importancia de estas capacidades y a sustantivarlas.

Más importante aún, **esta tesis contribuye a evidenciar la necesidad y la relevancia de abordar ambas cuestiones (la evaluación y los enfoques colaborativos) de forma conjunta para hablar de innovación en el diseño de políticas públicas**. Tal como apunta el apartado anterior, solamente cuando la evaluación tiene una voluntad real de aprendizaje y se enmarca en procesos de reflexión y de construcción de conocimiento colectivos (multi-sectoriales y multi-actor), es capaz de responder a contextos sociales complejos y cambiantes y proponer soluciones de política pública realmente adaptadas y efectivas.

5.3 Limitaciones de la investigación

Las principales limitaciones de esta tesis son las expuestas en cada uno de los artículos / capítulos. Sin embargo, más allá de estas cuestiones ya explicadas, es necesario apuntar una limitación adicional —en este caso de la tesis en su conjunto— derivada de que se realiza en el marco de un doctorado industrial. Más concretamente, del primer doctorado industrial en el que ha colaborado Ivàlua.

Ivàlua es una institución que trabaja por proyectos y la intención inicial, en el momento de iniciar esta tesis, era vincular cada uno de los artículos del compendio a un proyecto de Ivàlua. No obstante, por razones diversas, en la práctica resultó muy difícil conjugar esta realidad con la necesidad de relacionar los tres artículos con un único tema. Ello explica que no fuera posible plantear desde el inicio de la investigación cuál sería la estructura de la tesis —puesto que no era posible anticipar en qué proyectos trabajaría Ivàlua, y todavía menos yo— a uno, dos o tres años vista. También explica que la investigación terminara virando hacia un tema mucho más transversal como es la institucionalización de la evaluación y su relación con la innovación de políticas públicas; que no depende de proyectos concretos encargados a Ivàlua por las diferentes administraciones públicas, sino que está directamente relacionado con su misión institucional. En definitiva, esta realidad contextual ha condicionado la estructura de la tesis, especialmente la relación entre los diferentes capítulos/artículos.

5.4 Futuras líneas de investigación

Para profundizar en la relación entre evaluación e innovación sería importante abordar dos cuestiones que esta tesis no ha alcanzado a explorar: el papel tanto de los llamados *policy labs*, como de la experimentación. Los ***policy labs***, entendidos como centros dedicados principalmente a “producir programas y proyectos que buscan explorar ideas, resolver problemas, capacitar a líderes y proporcionar herramientas para mejorar los servicios públicos a partir de innovación” (Olejniczak et al., 2020:99) se han multiplicado en los últimos años, alrededor del mundo (Kim et al., 2023; Olejniczak et al., 2020). Aún y así, los escasos estudios empíricos comparados existentes muestran que, por ahora, los logros de los *policy labs* tienen un alcance limitado: por lo que respecta a la política pública, logran mejoras de carácter procesual —es decir, relacionadas con la implementación de las políticas— pero raramente cambios estructurales (Olejniczak et al., 2020); por lo que respecta a la capacidad de las administraciones públicas para desarrollar políticas públicas (*policy capacity*), los cambios en términos de capacidades analíticas y operacionales son notorios —tanto a nivel individual, como organizacional y sistémico— pero el desarrollo de capacidades políticas sigue siendo un desafío pendiente. En definitiva, la relación entre evaluación e innovación continúa siendo un desafío, no tanto a nivel teórico o académico sino más bien a nivel práctico.

El papel que puede jugar la **experimentación** no es menos controvertido. Si bien la idea de experimentar y de testear hipótesis se encuentra en el centro de la propuesta de “*policymaking* inteligente” de Sanderson (2009), los experimentos —entendidos de forma estricta como proyectos piloto evaluados a partir de métodos experimentales— no solamente están en el origen, sino que además son la base del movimiento de políticas basadas en evidencia (Baron, 2018; Cairney & Oliver, 2017). De hecho, la idea de “pirámide de la evidencia”, donde los métodos experimentales (los famosos RCTs, *randomised control trials*) están en la cúspide por ser considerados la forma más rigurosa de producción de evidencia continúa muy presente en buena parte de las organizaciones dedicadas a la evaluación, incluida Ivalua, como muestran los resultados del Capítulo 3 (Artículo 2). Por lo tanto, la cuestión está en si —tal vez en el marco de los mencionados *policy labs*— es posible ampliar el concepto de experimentación más allá de los métodos experimentales para ganar adaptabilidad sin perder rigor científico (aunque ello implique revisar cómo se define el rigor científico).

De hecho, sería interesante profundizar en el concepto de “**adaptabilidad**” de las políticas públicas, más allá de la experimentación. Tal como se ha comentado anteriormente, Head (2022:126) apunta que la búsqueda de nuevas soluciones debe combinar procesos inclusivos, evidencia rigurosa y adaptabilidad, definida como “un ajuste iterativo en la implementación de los programas”. Esta tesis ha profundizado sobre las dos primeras estrategias, pero la tercera sólo se ha abordado tangencialmente. Si bien Sanderson (2009) vincula esa capacidad de iterar a la evaluación, está claro que es un aspecto mucho más relacionado con la implementación de las políticas públicas. Por lo tanto, valdría la pena concretar en qué se traduce esa “adaptabilidad” en la gestión diaria de la política pública y cómo se relaciona con los enfoques colaborativos y con la innovación.

Por último, es relevante explorar qué otras vías o estrategias existen para innovar en el diseño de políticas públicas y responder mejor a la complejidad de los problemas

sociales, así como la relación que mantienen (o no) con la evaluación y los enfoques colaborativos. En ese sentido, Head (2022) identifica la **prospectiva** (*strategic foresight*) como una herramienta valiosa para responder de forma efectiva a retos cambiantes. Sería interesante profundizar sobre su relación tanto con la evaluación como con los enfoques colaborativos para identificar, también, complementariedades y sinergias.

En cualquier caso, es importante garantizar que, aunque todos estos análisis se desarrollen a nivel académico, y aunque buena parte de las aportaciones sean de carácter conceptual o teórico —como sería de esperar—, se busquen mecanismos para **acercar los resultados y recomendaciones a los policymakers**, pues al final son quienes pueden cambiar la práctica de las administraciones públicas. En ese sentido, vale la pena tomar en consideración —como apunta Wegrich (2019) respecto a las estrategias de innovación colaborativa— que los sesgos organizacionales y procedimentales de las administraciones públicas no son solamente una cuestión burocrática, sino que a menudo responden a una competición entre departamentos. De forma que, cualquier propuesta concreta de reforma debería tomar en consideración los mecanismos concretos que explican determinadas prácticas administrativas.

5.5 Referencias

- Ansell, C., & Gash, A. (2018). Collaborative Platforms as a Governance Strategy. *Journal of Public Administration Research and Theory*, 28(1), 16-32. <https://doi.org/10.1093/jopart/mux030>
- Baron, J. (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>
- Cairney, P. (2019). The UK government's imaginative use of evidence to make policy. *British Politics*, 14(1), 1-22. <https://doi.org/10.1057/s41293-017-0068-2>
- Cairney, P., & Oliver, K. (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>
- de Marchi, G., Lucertini, G., & Tsoukias, A. (2016). From evidence-based policy making to policy analytics. *Annals of Operations Research*, 236(1), 15-38. <https://doi.org/10.1007/s10479-014-1578-6>
- Head, B. W. (2022). Policy Innovation in Turbulent Times. En B. W. Head (Ed.), *Wicked Problems in Public Policy: Understanding and Responding to Complex Challenges* (pp. 123-139). Springer International Publishing. https://doi.org/10.1007/978-3-030-94580-0_7
- Jann, W., & Wegrich, K. (2007). Theories of the Policy Cycle. En F. Fischer & G. Miller (Eds.), *Handbook of Public Policy Analysis. Theory, Politics, and Methods*. Routledge. https://www.researchgate.net/publication/270876927_4_Theories_of_the_Policy_Cycle

Kim, S., Wellstead, A. M., & Heikkila, T. (2023). Policy capacity and rise of data-based policy innovation labs. *Review of Policy Research*, 40(3), 341-362.

Mulgan, G. (2005). *Government, knowledge and the business of policy making: The potential and limits of evidence-based policy*. <https://doi.org/10.1332/1744264053730789>

Newman, J., & Head, B. W. (2017). Wicked tendencies in policy problems: Rethinking the distinction between social and technical problems. *Policy and Society*, 36(3), 414-429. <https://doi.org/10.1080/14494035.2017.1361635>

Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widla, A., & Park, Y. (2020). *Policy labs: The next frontier of policy design and evaluation?* <https://doi.org/10.1332/030557319X15579230420108>

Parsons, W. (2002). From Muddling Through to Muddling Up- Evidence Based Policy Making and the Modernisation of British Government. *Public Policy and Administration*. <https://www.semanticscholar.org/paper/Making-and-the-Modernisation-of-British-Government-Parsons/3542aeae9824acc5207c9d581242b7ce429ff61c>

Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. SAGE Publications Ltd. <https://doi.org/10.4135/9781473913820>

Sanderson, I. (2002). Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration*, 80(1), 1-22. <https://doi.org/10.1111/1467-9299.00292>

Sanderson, I. (2009). Intelligent Policy Making for a Complex World: Pragmatism, Evidence and Learning. *Political Studies*, 57(4), 699-719. <https://doi.org/10.1111/j.1467-9248.2009.00791.x>

Wegrich, K. (2019). The blind spots of collaborative innovation. *Public Management Review*, 21(1), 12-20. <https://doi.org/10.1080/14719037.2018.1433311>