MINISTERIO DE INDUSTRIA, COMERCIO Y TURISMO

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

CUADERNOS ECONÓMICOS

Número 102 • 2021/II

La evaluación de las políticas públicas en España: recursos, metodologías y algunos ejemplos

Public policy evaluation in Spain: resources, methodologies and some examples

2021 CALIDAD REVISTAS CIENTÍFICAS ESPAÑOLAS

FECYT-457/2021
Fecha de certificación: 30 de julio de 2021 (7ª convocatoria)
Válido hasta: 30 de julio de 2022

EDITORA
María del Carmen Díaz Roldán
Universidad de Castilla-La Mancha

CONSEJO CIENTÍFICO
Michelle Boldrin, Washington University in St Louis
Mariam Camarero Olivas, Universitat Jaume I
Juan José Dolado, U. Carlos III de Madrid
Álvaro Escribano, U. Carlos III de Madrid
María Paz Espinosa, U. del País Vasco
Jesús Gonzalo Muñoz, U. Carlos III de Madrid
Andrew Jones, University of York
Timothy J. Kehoe, University of Minnesota
José María Labeaga, UNED
Inés Macho Stadler, U. Autònoma de Barcelona
Inmaculada Martínez Zarzoso, Universitat Jaume I
Matilde Mas Ivars, Universitat de València
Franco Peracchi, Georgetown University
Marta Rey García, Universidade da Coruña
Joan Enric Ricart, IESE, U. de Navarra

CONSEJO EDITORIAL
Carmen Arguedas Tomás, U. Autónoma de Madrid
Luis Ayala Cañón, U. Rey Juan Carlos
Hugo Benítez-Silva, Stony Brook University N.Y.
Florentino Felgueroso, U. de Oviedo
Ángel López Nicolás, U. Politécnica de Cartagena
Ana Lozano Vivas, U. de Málaga
Javier Salinas Jiménez, U. Autónoma de Madrid
Javier Suárez Pandiello, U. de Oviedo

DIRECCIÓN EJECUTIVA
Bernardo Hernández San Juan

JEFE DE REDACCIÓN
Antonio Alonso González

REDACCIÓN
Ana Belén Rodríguez Rodríguez

EDICIÓN Y REDACCIÓN
Ministerio de Industria, Comercio y Turismo
Secretaría de Estado de Comercio
Paseo de la Castellana, 162. Planta 12. 28046 Madrid
Teléfonos 91 349 6053 / 91 349 2548

DISTRIBUCIÓN
Ministerio de Industria, Comercio y Turismo
Centro de Publicaciones
Calle Panamá, 1. Planta 0, despacho 3. 28036 Madrid
Teléfonos: 91 349 5129 / 91 349 5133 / 91 349 4335
CentroPublicaciones@mincotur.es
www.revistasice.com

Esta publicación se encuentra en las siguientes bases de datos bibliográficas:
REBIUN, ECONLIT, DIALNET, LATINDEX, DICE, InDICES CSIC e ISOC.
Sus índices de impacto aparecen en IN-RECS.

Los análisis, opiniones y conclusiones expuestos en los artículos de esta publicación son los de los autores y no representan opiniones oficiales de la Secretaría de Estado de Comercio, con las que no tienen por qué coincidir necesariamente.

Editor: S.G. de Estudios y Evaluación de Instrumentos de Política Comercial. Secretaría de Estado de Comercio
Diseño de cubierta: TexTec
Composición y maquetación: TexTec
Impresión y encuadernación: Centro de Impresión Digital y Diseño de la Secretaría de Estado de Comercio
Papel:
Exterior: estucado mate ecológico (250 g)
Interior: offset ecológico FSC/TCF (80 g)
ECPMINCOTUR: 1.ª ed./200/1121
PVP: 18 € + IVA
DL: M 22392-1977
NIPO: 112-19-008-4
e-NIPO: 112-19-009-X
ISSN: 0210-2633
e-ISSN: 2340-9037

Catálogo general de publicaciones oficiales:
https://cpage.mpr.gob.es/

CUADERNOS ECONÓMICOS DE ICE

MINISTERIO DE INDUSTRIA, COMERCIO Y TURISMO

NÚMERO 102 • 2021/II

SUMARIO

• Presentación	Juan F. Jimeno	5
Visiones desde las Administraciones públicas		
• Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021)	José María Casado y Eloísa del Pino	13
• Los retos de la explotación masiva de datos administrativos para la investigación y las políticas públicas: una visión desde dentro de la Administración	Israel Arroyo y José Fernández Albertos	39
• Estadística oficial y evaluación de políticas públicas: oportunidades y retos	Juan Manuel Rodríguez Póo y Antonio Salcedo Galiano	51
Recursos y metodologías		
• Investigación económica y datos masivos: mercados, fines sociales y colaboración público-privada	José García Montalvo	69
• La evaluación como proceso de aprendizaje: una necesidad del presente, una inversión de futuro	Sara de la Rica Goiricelaya, Lucía Gorjón García e Imanol Lizarraga Álvarez	87
• La gestión de los datos administrativos en España: diagnóstico y retos de futuro	Miguel Almunia y Pedro Rey-Biel	103
Algunos ejemplos de evaluaciones de políticas públicas		
• La evaluación de impacto de las políticas activas de empleo en España: el deseo y la realidad	Inmaculada Cebrián López, Begoña Cueto Iglesias y Francisco Javier Mato Díaz	133
• Evaluación de impacto de políticas activas de empleo para colectivos de difícil inserción laboral	Yolanda Fátima Rebollo-Sanz y José Ignacio García Pérez	157
• El perfilado estadístico como instrumento para la evaluación del impacto del programa Incorpora	Jorge Casanova, Florentino Felgueroso, José Ignacio García Pérez y Sergi Jiménez-Martín	189
• Medidas educativas de refuerzo en lectoescritura: el caso del programa de Ayuda a la Dislexia en la Comunidad de Madrid.....	Almudena Sevilla, Pilar Cuevas-Ruiz, Luz Rello e Ismael Sanz	221

Presentación
La evaluación de las políticas públicas en España: recursos, metodologías y algunos ejemplos

Juan F. Jimeno
Banco de España, Universidad de Alcalá, CEMFI, CEPR e IZA

La evaluación de las políticas públicas en España sigue siendo una asignatura pendiente. Vivimos una época en la que abundan las bases de datos, en las que casi cualquier actividad queda registrada digitalmente, y en la que se dispone de instrumentos metodológicos, software informático y algoritmos de inteligencia artificial que permiten el tratamiento extensivo y el análisis intensivo de los datos. Tanto en el sector público como en el privado, la observación cuantitativa de fenómenos económicos y sociales es cada vez más extensa y precisa.

Estos desarrollos en la obtención y tratamiento de los datos permiten, entre otras cosas, que la evaluación de las políticas públicas pueda realizarse con mucho mayor rigor analítico y contenido empírico que en el pasado. Sin embargo, los investigadores sociales no siempre han tenido un acceso fácil a los datos necesarios para llevar a cabo estas tareas. En parte, la falta de cultura de evaluación económica en el pasado se traduce en que el diseño de las bases de datos no siempre se ha realizado teniendo en cuenta las necesidades de la evaluación. Además, los registros administrativos donde se depositan los datos tienen otras finalidades relativas a la gestión de servicios públicos que no siempre casan bien con esas necesidades. Por otra parte, las bases de datos necesarias para realizar un determinado ejercicio de evaluación de políticas públicas a menudo se encuentran dispersas entre distintos organismos de las Administraciones públicas, que no se muestran siempre propicios a compartir esos datos para la creación de una muestra amplia de individuos y variables con las que realizar dichas evaluaciones. Como resultado, muchas de las medidas de políticas económicas y sociales no son evaluadas suficientemente y, por consiguiente, se siguen diseñando y ejecutando sin la información ni el conocimiento suficientes que garanticen su efectividad.

Este número 102 de *Cuadernos Económicos de Información Comercial Española* tiene como razón de ser proporcionar una panorámica actual de la evaluación de políticas públicas en España. Para ello se han seleccionado diez artículos estructurados en tres apartados. En el primero de ellos («Visiones desde las Administraciones públicas»), que contiene tres artículos, se presenta el punto de vista de investigadores, que actualmente ocupan responsabilidades en las Administraciones públicas que recopilan y usan gran cantidad de datos económicos y sociales, acerca de la administración, uso y puesta a disposición de investigadores externos de los datos de registros administrativos o encuestas oficiales. En el segundo apartado («Recursos y

metodologías») que también consta de tres artículos, investigadores académicos con amplia experiencia en la evaluación de políticas públicas con diversas fuentes de datos ofrecen su valoración del estado de la evaluación de políticas públicas en España. En el apartado final («Algunos ejemplos de evaluaciones de políticas públicas») se presentan cuatro aplicaciones, tres relativas a políticas de empleo, un campo donde la evaluación se ha desarrollado enormemente, y una de un programa educativo, ámbito en el que la evaluación también es cada vez más frecuente.

En el primer artículo **José María Casado** y **Eloísa del Pino**, autores afiliados a la Autoridad Independiente de Responsabilidad Fiscal (AIReF), presentan una panorámica de la evaluación de políticas públicas por las administraciones central y autonómicas españolas durante las dos últimas décadas. Prestando atención al marco normativo y organizativo, la capacidad evaluadora de las instituciones públicas y la práctica y el uso de la evaluación, destacan las características y los márgenes de mejora en la evaluación de las políticas públicas en estos ámbitos con el objetivo de garantizar la independencia, la investigación sistemática y la relevancia analítica.

La propia creación de la AIReF, la extensión de sus responsabilidades a ejercicios de evaluación de políticas de gasto público (los llamados *Spending Reviews*) y la reciente creación de una División de Evaluación del Gasto Público en su seno han sido hitos recientes en el proceso de sistematización y profesionalización de la evaluación de políticas públicas dentro de las Administraciones públicas en nuestro país. Según Casado y del Pino, si bien la estructura institucional de evaluación se ha extendido notablemente, todavía quedan aspectos por desarrollar, tales como una formalización legal que garantice continuidad y regularidad temporal en las evaluaciones, reglas que impulsen la independencia de las agencias evaluadoras y flexibilidad en la contratación para que estas puedan disponer de los recursos necesarios para cumplir sus funciones con profesionalidad e investigación analítica y sistemática. Finalmente, concluyen que la realización de evaluaciones por las Administraciones públicas es necesaria pero no resultaría suficiente si no se divulgaran sus resultados adecuadamente ni se incluyeran en la toma de decisiones sobre políticas públicas.

En el segundo artículo, **Israel Arroyo** y **José Fernández Albertos**, del Ministerio de Inclusión, Seguridad Social y Migraciones, se centran en el papel de las Administraciones públicas como creadoras y gestoras de registros administrativos que proporcionan los datos que constituyen la principal materia prima para muchos ejercicios de evaluación. Para el cumplimiento de sus funciones, las Administraciones públicas recogen rutinaria y regularmente gran cantidad de datos. Las capacidades de almacenamiento y de tratamiento estadístico de esos datos han aumentado espectacularmente y permiten ahora análisis estadísticos y económicos mucho más precisos y detallados. Además, su amplitud en el espacio y en el tiempo permite que la validez interna y externa de esos estudios sea muy superior a la que se puede obtener con otros tipos de datos.

No obstante, como señalan Arroyo y Fernández Albertos: «la puesta a disposición de estas bases de datos para finalidades académicas y de evaluación… no es sin embargo un proceso sencillo». Con la experiencia y la perspectiva de la que

disponen por su filiación con uno de los departamentos ministeriales que más propicio se ha mostrado a facilitar el uso de datos administrativos por investigadores externos, presentan, en primer lugar, una taxonomía de las dificultades (marco legal garantista en aras de la protección de la confidencialidad y la privacidad, la ausencia de conocimiento y experiencia en materia de protección de datos en el seno de las Administraciones públicas, la fragmentación y dispersión de competencias y la escasa predisposición a favor de la evaluación). En segundo lugar, discuten los posibles cambios del marco legal en aras del uso de datos administrativos para la investigación y la evaluación con la suficiente seguridad jurídica de las instituciones públicas que facilitan los datos, así como la necesidad de desarrollar infraestructuras estables para la cesión y explotación de datos.

El primer apartado de este número concluye con un artículo de **Juan Manuel Rodríguez Póo** y **Antonio Salcedo Galiano**, ambos afiliados al Instituto Nacional de Estadística (INE), sobre el uso de datos administrativos y nuevos instrumentos en la elaboración de las estadísticas oficiales. Si bien las estadísticas oficiales siempre están en un continuo proceso de revisión, con la pandemia de la COVID-19 se ha puesto de manifiesto la necesidad de mejorar la eficiencia en la recogida y tratamiento de los datos para así aumentar la frecuencia, granularidad y calidad de los resultados que se ofrecen a la opinión pública.

Rodríguez Póo y Salcedo discuten, en primer lugar, la utilización de indicadores socioeconómicos en los ejercicios de evaluación de políticas públicas y cómo pueden construirse «contrafactuales» que permitan medir relaciones causa-efecto entre una determinada medida y sus objetivos. Tras repasar la evolución de la construcción de estadísticas públicas en España y en el contexto europeo y las disposiciones relativas al acceso a los microdatos para análisis estadísticos de fines científicos, presentan algunos desarrollos recientes de estadísticas experimentales y destacan la importancia de que, tanto en el ámbito nacional como en el europeo, el sistema estadístico pueda desarrollar funciones de administrador de datos y participar en el diseño de estructuras de gobernanza de los datos que facilite el acceso a los mismos, tanto a las distintas agencias públicas como a investigadores externos con fines estadísticos o científicos.

El segundo apartado de este número que contiene contribuciones de investigadores externos sobre la situación de la disponibilidad de datos y la evaluación de políticas públicas, comienza con un artículo de **José García Montalvo** sobre las perspectivas que la disponibilidad de datos masivos ofrece para la investigación económica, prestando especial atención a experiencias recientes de colaboración entre los sectores público y privado y a la aplicación de técnicas de aprendizaje automático (*machine learning*) al tratamiento de datos masivos (*big data*). Tras repasar el desarrollo de estas bases de datos y los nuevos instrumentos de análisis en el sector privado, se ofrece una panorámica de su creciente introducción en la investigación social y económica, destacando las ventajas del *big data* en términos de amplitud en el espacio (abarcan toda la población), en el tiempo (son longitudinales) y de precisión (menores errores de medida, de imputación o de pérdida de muestra),

presentando ejemplos de trabajos de investigación recientes en varios campos de la Economía que han utilizado *big data*.

Un aspecto muy relevante en la extensión de la disponibilidad rápida de grandes bases de datos tiene que ver con la mayor participación del sector privado en la generación y gestión de grandes bases de datos, lo que abre oportunidades de colaboración público-privada a la hora de facilitar la disponibilidad de grandes bases de datos de alta frecuencia que permitan un seguimiento rápido y preciso de la situación socioeconómica. García Montalvo describe algunos ejemplos de esa colaboración en varios países, entre ellos España. En sus conclusiones destaca que, si bien hasta hace pocos años España se situaba en una posición retrasada en el contexto internacional en lo que respecta al aprovechamiento de datos administrativos de instituciones públicas y a la colaboración público-privada para la creación y gestión de grandes bases de datos, en los últimos tiempos hay varias iniciativas que sugieren una mayor sensibilización de las instituciones públicas a este respecto. No obstante, en su opinión todavía queda mucho camino por recorrer para que la opinión pública y los medios de comunicación aprecien la importancia de esta información y de su uso para la investigación económica y social.

El segundo artículo del apartado del monográfico, que versa sobre «Recursos y Metodologías», **Sara de la Rica, Lucía Gorjón** e **Imanol Lizarraga**, abogan por una «nueva cultura de la evaluación de políticas públicas» que contribuya a la mejora de las administraciones y a la resolución de los retos socioeconómicos que afrontamos. Lo hacen, primero, apelando a la realización de una «evaluación integral» que contemple todas las fases del programa en cuestión, desde su fase de diseño hasta la presentación y divulgación de los resultados; en segundo lugar, resaltando la importancia de la generación, gestión y uso de datos administrativos en la evaluación de políticas públicas y, finalmente, presentando brevemente un ejemplo de evaluación del contrato laboral en prácticas realizado por ellos mismos.

En sus conclusiones, de la Rica, Gorjón y Lizarraga destacan que en un mundo con más y mejores fuentes de información y en el que la complejidad de los problemas socioeconómicos se agudiza, la mayor riqueza de las administraciones e instituciones públicas son los datos que generan y que poseen. Sin ellos y sin la evaluación integral de las políticas públicas no será posible mejorar la eficacia y la eficiencia con la que se asignan los recursos públicos y se diseñan dichas políticas. Desaprovechar esta riqueza es un lujo que no nos podemos permitir.

El segundo apartado del número concluye con un artículo de **Miguel Almunia** y **Pedro Rey Biel**, sobre la situación actual en España y los retos de futuro de la gestión de datos administrativos. Su punto de partida es que, en comparación con otros países de nuestro entorno, España se está quedando rezagada. Su aproximación a esta conclusión comienza con una descripción del marco conceptual para la gestión de los datos administrativos en la que se discuten las principales cuestiones a tener en cuenta: quién debe custodiar los datos, cuáles deben ser los criterios de acceso, si el acceso debe ser gratuito, si los organismos que proporcionan los datos deberían revisar los resultados de las investigaciones que se hacen con ellos, y cuáles son las

plataformas (salas seguras o conexión remota) más adecuadas para acceder a los datos. A continuación, pasan revista a cómo se han resuelto estas cuestiones en los países nórdicos, Francia, Alemania, Portugal, Estados Unidos y Reino Unido, y cuál es la situación actual en España. Finalmente, realizan algunas propuestas para que la comunidad científica tenga acceso, de manera controlada y con todas las garantías de confidencialidad y privacidad de la población, a registros administrativos.

En sus conclusiones Almunia y Rey Biel destacan que se necesita un cambio de modelo en la gestión de datos administrativos en la fragmentada Administración pública española que permita su mejor coordinación y uso. Aunque hay casos recientes de fracaso, sobre todo con la gestión de la información sobre la incidencia de la pandemia de la COVID-19, se muestran esperanzados en que otros casos aislados de éxito y de iniciativas conjuntas de varias instituciones puedan liderar el cambio de modelo. Señalan que un hito importante en este camino sería la articulación de un Plan Nacional que impulsara la creación de una agencia de datos.

El último apartado de este número presenta algunos ejercicios de evaluación de políticas públicas referidos a dos de los campos en los que estos ejercicios son relativamente más frecuentes: el mercado de trabajo y el sector educativo. En primer lugar, **Inmaculada Cebrián, Begoña Cueto** y **Francisco Javier Mato** analizan la evaluación de impacto de las políticas activas del mercado de trabajo en España. Tras describir y analizar el marco legislativo que regula las políticas activas del mercado de trabajo y su evaluación y el Fondo Social Europeo y su evaluación, avanzan una serie de propuestas para mejorar la gestión de esas políticas y contribuir a hacerlas más efectivas.

En su análisis destacan que el marco regulatorio de la evaluación de políticas activas del mercado de trabajo todavía denota una falta de comprensión de lo que significa una evaluación de impacto que aspire a establecer una relación causal entre participación en un determinado programa de empleo y cambios en la situación laboral de los participantes. En cuanto a la evaluación de los programas del Fondo Social Europeo, destacan que, por ser descentralizada, ha permitido que algunas administraciones hayan superado los mínimos exigidos por la Unión Europea, con evaluaciones de impacto que pueden contribuir a favorecer la difusión de buenas prácticas de evaluación. En cuanto a las principales barreras a la mejora de la evaluación de las políticas activas del mercado de trabajo en España, señalan la falta de colaboración entre las administraciones que son propietarias de la información y las entidades que podrían llevar a cabo evaluaciones de impacto, la existencia de datos y resultados que no se publican ni divulgan suficientemente y la falta de homogeneidad entre las administraciones a la hora de abordar la evaluación de dichas políticas.

En el segundo ejemplo de ejercicios de evaluación, **José Ignacio García Pérez** y **Yolanda F. Rebollo Sanz** presentan evaluación de impacto de dos medidas de una política activa del mercado de trabajo dirigida a parados de larga duración, tanto jóvenes como adultos, desarrolladas en Andalucía durante los años 2016-2018: los programas Emple@Joven y Emple@30+, que ofrecen experiencia laboral durante periodos de entre 3 y 6 meses. Con los datos disponibles y las técnicas

cuantitativas habituales, concluyen que los programas solo fueron efectivos para colectivos muy determinados con especiales problemas de inserción laboral, lo que sugiere que para aumentar la efectividad de esos programas sería necesario un ejercicio *ex ante* de perfilado estadístico que permitiera que los programas fueran más selectivos en cuanto a la población objetivo.

Además de mostrar cómo se pueden llevar a cabo ejercicios de evaluación de impacto de iniciativas locales de empleo cuando los datos se ponen a disposición de investigadores externos y las agencias implicadas se muestran interesadas en los resultados, García Pérez y Rebollo llevan a cabo una excelente presentación de cómo se pueden identificar los efectos de una determinada medida mediante la construcción rigurosa de grupos de control y de tratamiento con los datos habitualmente recogidos por las agencias responsables y de cómo interpretar los resultados de las comparaciones entre ambos grupos. Los resultados, aunque no sean especialmente positivos en relación con la empleabilidad de los participantes de los programas analizados, deberían contribuir a difundir buenas prácticas tanto en lo relativo al diseño de programas de políticas activas del mercado de trabajo como a su evaluación.

También en relación con la evaluación de las políticas activas del mercado de trabajo, **Jorge Casanova, Florentino Felgueroso**, **José Ignacio García Pérez** y **Sergi Jiménez Martín** muestran como en el contexto de un ejercicio de evaluación se puede desarrollar una metodología de perfilado estadístico que ayude a mejorar la efectividad futura del programa en cuestión. Lo hacen utilizando el programa Incorpora que, además, tiene el interés de haber sido desarrollado fuera del sector público (por la Fundación "La Caixa") y que consiste en mecanismos de tipo formativo para apoyar la empleabilidad y recolocar en el mercado de trabajo a personas de difícil inserción laboral y en riesgo de exclusión. Según sus resultados, el programa aumentó la empleabilidad media de los participantes, si bien con una intensidad decreciente en la duración de la situación de desempleo de los mismos.

Es este tipo de ejercicios el que mejor demuestra el potencial de la evaluación de impacto. Con la mayor disponibilidad de datos, el interés creciente de las agencias responsables de políticas activas del mercado de trabajo por conocer sus efectos y mejorar su efectividad, la creciente colaboración público-privada y la disponibilidad de investigadores externos para acometer este tipo de ejercicios a un coste muy reducido, dicho potencial es considerable. Como muestran Casanova, Felgueroso, García Pérez y Jiménez Martín, los propios ejercicios de evaluación permiten desarrollar nuevas herramientas y metodologías que guíen el diseño y la implementación de futuros programas más efectivos.

Finalmente, el número se cierra con un ejercicio de evaluación de un programa educativo implementado en la Comunidad de Madrid en el curso 2018-2019 de Ayuda a la Dislexia con el objetivo de reforzar las capacidades de lectoescritura de alumnos de primaria con ese tipo de dificultades realizado por **Almudena Sevilla, Pilar Cuevas Ruiz, Luz Rello** e **Ismael Sanz.** Además de medir los efectos de la participación del centro educativo en el programa sobre el rendimiento de sus alumnos en pruebas externas y estandarizadas de matemáticas, lengua e inglés, tratan de estimar

el impacto del estatus socioeconómico de los alumnos en la relación causa-efecto de interés, proporcionando una primera aproximación a este tipo de programas educativos que permitirán una evaluación más precisa de nuevas convocatorias.

Este tipo de programas tiene dos fases: la identificación de los alumnos con dificultades (la selección del grupo de tratamiento) y la ejecución de los instrumentos para combatirlas. En ambas, la evaluación permite obtener resultados interesantes. Como muestran Sevilla, Cuevas Ruiz, Rello y Sanz, la identificación de los alumnos con dificultades lectoescritoras debe hacerse con consideración a las diferencias de género y también posiblemente de otras características socioeconómicas. Igualmente muestran que el diseño y la ejecución de los programas a veces dificultan la propia evaluación, por lo que esta debe considerarse una parte integral del programa desde la etapa inicial del mismo.

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021)

Developments and challenges of public policy evaluation in the Spanish administration (2000-2021)

José María Casado
Eloísa del Pino
AIReF

Resumen

En este artículo se analiza la evolución de la evaluación de políticas públicas en las Administraciones central y autonómica españolas en las últimas dos décadas. En primer lugar, se explora la existencia en el caso de España de los elementos existentes en los países de nuestro entorno que más han avanzado en el proceso de institucionalización de la evaluación (marco normativo y organizativo, capacidad evaluadora y actitudes afines a la evaluación, y la práctica y uso de la evaluación). También se discute hasta qué punto los elementos que configuran nuestro sistema de evaluación logran garantizar los principios de independencia, investigación sistemática, profesionalidad y relevancia de los análisis que la literatura sobre institucionalización considera esenciales para el buen funcionamiento de los sistemas de evaluación.

Palabras clave: *evaluación de políticas públicas, institucionalización, principios de la evaluación, España.*

Clasificación JEL: *C00, D6, D7, H1, H8.*

Abstract

This article analyses the evolution of public policy evaluation in Spain between 2000 and 2021, at the central and regional levels. To do so, we first assess the presence in Spain of the elements that are present in countries that have made progress in the process of institutionalizing policy evaluation (a regulatory and organizational framework that supports evaluation, evaluation capacity and whether attitudes towards evaluation among policy makers, and whether evaluations are practiced and used). Secondly, we discuss the extent to which the elements that make up the Spanish evaluation system manage to guarantee the core principles (independence, systematic enquiry, professionalism and relevance of analysis) that the literature on institutionalization considers essential for the proper functioning of evaluation systems.

Keywords: *public policy evaluation, institutionalization, core evaluation principles, Spain.*

DOI: https://doi.org/10.32796/cice.2021.102.7308
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

Los continuos avances y retrocesos experimentados en materia de evaluación en las últimas décadas en España han instalado entre los observadores cierto pesimismo, o al menos escepticismo, sobre la posibilidad de que las Administraciones españolas sean capaces de examinar de forma crítica y objetiva el impacto de la acción pública, repensar sus políticas y rendir cuentas a la ciudadanía. Desde entonces, se ha argumentado que nuestra Administración carece de los incentivos y de la capacidad interna suficiente para hacer evaluación de políticas públicas (en adelante, EVPP) o que no tiene cultura evaluadora. También se ha argumentado que, a pesar de que dispone de una gran cantidad de datos para hacer evaluación, es muy difícil acceder a ellos o que, cuando la evaluación se lleva a cabo, no se hace con el enfoque metodológico adecuado o no tiene la suficiente objetividad o trascendencia o, finalmente, que no se utiliza para mejorar las políticas y la eficacia y eficiencia del gasto público (Ruiz, 2012; Garde-Roca, 2006, 2017; De la Rica, 2020; Munárriz, 2020; FEDEA, 2021).

A pesar de esta percepción sobre las supuestas carencias de nuestro sistema, que teóricamente impedirían la consolidación de un sistema de EVPP en España, y, sin dudar de que la evaluación presenta todavía importantes desafíos, durante los últimos años, los distintos niveles de gobierno han puesto en marcha numerosas iniciativas y se ha desarrollado cierta actividad evaluadora.

En este artículo nos preguntamos cuál es la situación actual de la EVPP en España: hasta qué punto está justificada la preocupación de los expertos; en qué medida nuestro sistema se aleja del existente en otros países de nuestro entorno y de los elementos básicos que los especialistas identifican como necesarios para hablar de un sistema consolidado; y, finalmente, cuáles son los principales retos que afronta la EVPP para asegurarse su independencia, credibilidad y trascendencia.

Para responder a esta pregunta se propone el concepto de institucionalización de la EVPP y se analizan sus dimensiones para el caso de España. Se entiende por institucionalización «el proceso de incorporación de normas y rutinas al sistema institucional existente con el fin de estimular la demanda de evaluaciones por parte de la sociedad civil y de las propias autoridades y de garantizar tanto su práctica sistemática, independiente y profesional como su utilización por los responsables públicos en la adopción de decisiones» (AIReF, 2021, p. 28).

El concepto o, en realidad, los distintos conceptos de institucionalización han sido aplicados tanto para tratar de medir o analizar la situación de la EVPP en numerosos países en el mundo, incluida España, como para estimar en qué dimensiones es necesario fortalecer el sistema de EVPP existente en un país o región (por ejemplo, Jacob, Speer & Furubo, 2015; Stockmann, Meyer & Taube, 2020a, 2020b; Bustelo, 2020; AIReF, 2021). Este artículo busca contribuir a la literatura sobre institucionalización de la EVPP, discutiendo críticamente el propio concepto y aplicándolo al caso de España con evidencia sobre las Administraciones central y autonómica y su evolución entre 2000 y 2021.

Como anticipo de las conclusiones, se puede sostener que, tras una etapa de incertidumbre, España ha realizado avances sustantivos en materia de evaluación en los últimos años. No obstante, y dado que nuestro país ya experimentó importantes retrocesos en el proceso de institucionalización de la EVPP, conviene consolidar estos avances antes de atreverse a confirmar que el sistema está ya maduro.

A partir de ahora, el artículo se organiza en otros tres epígrafes. En primer lugar, se revisa críticamente el concepto de institucionalización de la EVPP, se realiza una propuesta para su aplicación empírica, que incluye el análisis de los llamados principios de la evaluación. En segundo lugar, se repasan los avances y retrocesos del proceso de institucionalización en los niveles central y autonómico de gobierno en España para el periodo comprendido entre 2000 y 2021. Y finalmente, se discuten los hallazgos y los retos que afronta la EVPP en España.

2. La institucionalización y los principios de evaluación: definiciones, dimensiones, limitaciones y propuesta para su aplicación empírica

En las últimas décadas, la EVPP parece haber recibido un nuevo impulso en las democracias avanzadas. Junto a la presión que reciben los Gobiernos para mejorar la eficacia y eficiencia de crecientes volúmenes de gasto público y de profundizar en la calidad de las democracias a través de la transparencia, en los últimos años, el avance de las políticas informadas en la evidencia y el propio desarrollo de metodologías, tecnologías y la disponibilidad de datos, han impulsado la evaluación. Entendida como «la aplicación sistemática de procedimientos de las ciencias sociales para valorar la conceptualización, el diseño, la implementación y la utilidad de las intervenciones públicas» (Rossi, Freeman & Lipsey, 2003, p.5), mediante la EVPP, las autoridades tratan de adoptar mejores decisiones, ser más transparentes, rendir cuentas sobre la utilización de los recursos públicos y su impacto y, también, en ocasiones, legitimar sus decisiones.

Sin embargo, ni todos los países ni dentro de ellos todas las organizaciones públicas muestran el mismo «activismo evaluador» (Pattyn, 2012). En muchos lugares la EVPP es una práctica poco normalizada e incluso esporádica. A veces, se lleva a cabo solo en un sector o en un organismo o depende de la voluntad de responsables públicos concretos, que durante un tiempo o en circunstancias específicas impulsan este ejercicio. Algunos Gobiernos se comprometen a evaluar determinadas políticas, debido a que esta es una exigencia para recibir fondos proporcionados por terceros o porque precisan revisar importantes programas de gasto o, en ocasiones, programas piloto específicos.

Más allá de algunos factores generales –por ejemplo, el ser o no una democracia, haber avanzado hacia el paradigma de la Nueva Gestión Pública o el pertenecer y/o recibir financiación procedente de instituciones terceras– (Del Pino & Colino, 2021), no están claros los factores que contribuyen al surgimiento de la llamada *capacidad evaluadora*, entendida como «la habilidad de asegurar recursos materiales,

tecnología y profesionales para poder desarrollar evaluaciones creíbles e independientes de políticas, garantizar su utilización en la adopción de decisiones de política pública y en último término conseguir que se institucionalice» (AIReF, 2021).

Entre los expertos existe el convencimiento de que la calidad y la utilidad de la EVPP depende de la medida en que la capacidad evaluadora logre incardinarse en el engranaje institucional existente y perpetuarse en el tiempo, de modo que la práctica de la evaluación deje de ser un ejercicio aislado (Garde-Roca, 2006). Se pueden llevar a cabo mejores evaluaciones si, desde el origen de las políticas, se documentó el proceso de decisión identificando alternativas, si se especificó cuál era el estado ideal del problema al que se aspiraba, los objetivos a lograr, si se prepararon los sistemas de información o se establecieron los indicadores para medir la implementación y el impacto de la política. La trascendencia de la EVPP será mayor si está previsto que forme parte del proceso de rediseño de las políticas, si los evaluadores han acumulado suficiente experiencia para hacer participar a los evaluados e integrar sus necesidades y definir con ellos sus estrategias. Igualmente, la influencia de la EVPP será mayor si los responsables públicos se ven interpelados por la sociedad civil u otras autoridades para la utilización de sus resultados, con el fin de mejorar las intervenciones públicas.

Por ello, los expertos y numerosas organizaciones internacionales como Naciones Unidas (CEPAL, 2021), el Banco Mundial (Rist, Boily & Martin, 2011), la OCDE (2021) o la Unión Europea (2007), o Lázaro (2015) han reflexionado sobre cómo caminar hacia la institucionalización de la EVPP. En la actualización del *Atlas Internacional de la Evaluación* de Furubo, Rist y Sandahl (2002), Jacob, Speer y Furubo (2015, pp. 10-11) analizan la situación de la institucionalización en 19 países a los que puntúan en nueve dimensiones e indicadores.

Por su parte, Meyer, Stockmann y Taube (2020, p. 15), en su intento reciente de definir y estimar de manera más completa el grado de institucionalización de la EVPP, sugieren comprobar la existencia de diferentes dimensiones en tres sistemas[1]: el sistema político, el social y el profesional. En realidad, los autores formalizan la idea, presente en toda literatura sobre institucionalización, de que la consolidación de un sistema de evaluación de políticas implica no solo a las autoridades, sino también a la propia sociedad y a la comunidad profesional.

En su análisis de la evaluación en 17 países y en la Unión Europea, Stockmann, Meyer y Taube (2020a) pidieron a los expertos nacionales que analizan las características de diferentes dimensiones de la institucionalización, que trataran de identificar determinados indicios en los tres sistemas mencionados más arriba. En relación con el *sistema político* debían identificar en su país: 1) una política pública o estrategias de evaluación y de regulación general, o al menos sectorial, en la que se establezca la independencia, se regule la calidad de la evaluación, su orientación a medir el impacto, su presupuesto y se integre de algún modo en el Parlamento; 2) la práctica de la

[1] Véase la discusión sobre el concepto y su vinculación con la sociología, la economía y la ciencia política en Meyer, Stockmann y Taube (2020, p. 8 y siguientes).

evaluación, su frecuencia en diferentes sectores de política, los determinantes de la evaluación, el tipo de evaluación que se practica, su naturaleza interna o externa y si existe o no un organismo evaluador; y 3) hasta qué punto, quién, para qué y por qué se usa la evaluación, en qué sectores se utiliza más y con qué mecanismos se garantiza su uso.

Por lo que se refiere al *sistema profesional,* deberían identificarse: 1) formación de educación superior; 2) publicaciones científicas y técnicas, guías, plataformas de comunicación, organizaciones profesionales, un mercado para la evaluación y la profesión de evaluador; y 3) estándares y reglas desarrolladas por estos profesionales, exigidas por los propios clientes y seguidas por los evaluadores.

En lo que respecta al *sistema social,* se avanza en el proceso de institucionalización: 1) si es usual en el país que las evaluaciones se utilicen para informar las decisiones o los procesos participativos; si se utiliza por parte de la sociedad civil, individuos, empresas u organizaciones del tercer sector; si la sociedad participa en las evaluaciones; 2) si la sociedad conoce y entiende el significado de la evaluación, si los informes se publican y discuten en los medios; y 3) si existe una demanda de evaluación hacia los decisores públicos procedente de la sociedad civil.

Meyer, Stockmann y Taube (2020) no incluyen ninguna mención directa a la *cultura evaluadora* como tal, que otros autores consideran un ingrediente esencial para la institucionalización. Sin embargo, resulta plausible que para que la evaluación pueda finalmente institucionalizarse, sea necesario que la sociedad civil y los responsables públicos compartan ciertas actitudes hacia la misma. La cultura evaluadora podría definirse como «un patrón de creencias, valores y conocimientos básicos compartidos por los responsables de las distintas fases del ciclo de las políticas que les proporcionan normas de comportamiento en relación con la demanda, la práctica, la utilización y difusión de las evaluaciones y la promoción y participación de la sociedad en este proceso» (definición adaptada de De Peuter & Pattyn, 2009, p. 6).

A pesar de la larga enumeración de diferentes dimensiones que indicarían el grado de institucionalización de la EVPP, se sabe poco sobre qué elementos concretos hacen que los sistemas de evaluación sean exitosos y más bien parece que pueden encontrarse distintos modelos eficaces en distintos contextos. Por ejemplo, Rosentein (2015) examina lo que llama el estatus de la Política Nacional de Evaluación en 115 países. De ellos, solo 20 tienen una norma en la que se recoge la política de evaluación. Algunos realizan evaluación de modo rutinario, a pesar de no tener una norma. Al mismo tiempo, la evaluación estaba a cargo de diferentes tipos de organismos que pueden ser igualmente eficaces, a pesar de sus diferencias institucionales. En definitiva, ni la existencia de una norma ni la centralización de la evaluación en un solo organismo parecen elementos necesarios para la institucionalización de la EVPP.

Asimismo, un reciente estudio de la OCDE (2021), realizado a instancias de la AIReF[2], muestra que los países que han avanzado en la institucionalización de la evaluación comparten algunas similitudes respecto a sus sistemas de evaluación,

[2] Este estudio ha sido financiado y ha contado con la supervisión técnica de la Comisión Europea (DG REFORM).

pero también diferencias. En algunos lugares no se puede mencionar una agencia concreta que practique la EVPP. La evaluación se lleva a cabo por centros gestores de la propia Administración con funciones de evaluación y, en ocasiones, estas evaluaciones se coordinan por un organismo. En otros lugares, existen instituciones independientes evaluadoras. Además, los departamentos de Hacienda suelen jugar un papel relevante en lo que se refiere a la integración de los resultados de las evaluaciones en el ciclo presupuestario. Un elemento común destacable es la existencia de múltiples iniciativas para mejorar la disponibilidad de datos, la actualización periódica de los sistemas de información, su interconexión y el establecimiento de las condiciones para su utilización. También exhiben importantes diferencias respecto a la existencia de normas, el grado de centralización de estas tareas dentro de sus Gobiernos centrales o el papel de los Parlamentos en el sistema.

Dados estos hallazgos, más que tratar de identificar fórmulas normativas o fórmulas organizativas específicas, parece más útil pensar en cómo cada país, teniendo en cuenta sus peculiaridades político-administrativas y socioeconómicas, puede garantizar que su sistema de evaluación asegure una serie de principios que se predican de los sistemas de evaluación exitosos. Por ejemplo, el Banco Mundial (2019) ha sugerido que estos principios pueden resumirse en credibilidad, utilización e independencia. La propia AIReF (2021) ha enumerado los siguientes, que guían el desarrollo de su actividad evaluadora: la *investigación sistemática*, entendida como la realización de un análisis riguroso de datos relevantes, utilizando las metodologías más apropiadas que provean de evidencias objetivas y factuales a sus evaluaciones y sobre las que se sustenten las propuestas y recomendaciones; *la independencia*, definida como la ausencia de conflictos de interés entre los evaluadores y de presión política u organizativa, así como la autonomía en la realización de las evaluaciones y en la difusión de los resultados; *la profesionalidad de los equipos*, que poseerán los conocimientos y habilidades necesarios para evaluar y mantener la integridad del proceso, incluyendo en él la participación de los evaluados; *la relevancia o transcendencia* de los análisis, seleccionando programas a evaluar en función de su importancia social o presupuestaria y facilitando la utilización de sus resultados.

En resumen, para explorar hasta qué punto los sistemas de EVPP están institucionalizados, proponemos, en primer lugar, analizar el grado de desarrollo de una serie de elementos que aparecen en los sistemas de evaluación más reconocidos y que están relacionados: comprobar si existe un marco regulatorio y una estructura organizativa que pueda amparar el desarrollo de la evaluación, la capacidad evaluadora y valores afines a la evaluación; y, por último, si hay una actividad evaluadora sostenida y el fruto de las mismas se utiliza. En segundo lugar, proponemos analizar hasta qué punto estos elementos logran garantizar los principios de evaluación. A continuación, aplicamos esta propuesta al análisis del caso de España.

3. La evaluación de políticas en España a partir del año 2000

3.1. Antecedentes: de la pertenencia europea a la extinción de la AEVAL

Hasta los años 2000, España no inicia un proceso deliberado de construcción de un sistema de evaluación integral. Entre los años 1980 y 1990, las evaluaciones que se habían llevado a cabo se concentraron en determinados sectores de política pública (educación, empleo, servicios sociales, sanidad, fondos europeos o cooperación al desarrollo) y se realizaban de forma aislada (Fenstein & Zapico-Goñi, 2010; Bustelo, 2020). Como es bien conocido, España es uno de los países lentos en cuanto a la implantación de la Nueva Gestión Pública, debido a razones institucionales y de cultura político-administrativa (Pollitt & Bouckaert, 2017). Ello puede explicar cierto retraso respecto a otros países en el despegue de la evaluación, situación que empieza a corregirse a partir del ingreso en la UE, una de las razones de su puesta en marcha en los primeros años 2000.

3.1.1. El marco organizativo y normativo

En la primera década de 2000, el paso más significativo en materia de evaluación fue la creación de la Agencia Estatal de Evaluación de las Políticas Públicas y Calidad de los Servicios (AEVAL), en la Ley 28/2006 de Agencias Estatales. Se reguló mediante el RD 1418/2006, por el que se aprueba su Estatuto. La Agencia estaba vinculada al ejecutivo y su presidente era nombrado por el Gobierno. Los encargos de evaluación se recogían en Acuerdo del Consejo de Ministros a propuesta de la Agencia, que debía ser aprobada por el pleno de su Consejo Rector. Además, varias previsiones legales incluían mandatos de evaluación adicionales de programas recogidos en el Programa Nacional de Reformas y en la Estrategia de Desarrollo Sostenible.

Tras una primera década de crecimiento en la que realizó múltiples evaluaciones, la AEVAL fue perdiendo su impulso inicial. A partir de 2011, muchas de las demandas hacia la AEVAL estuvieron más relacionadas con la consultoría de gestión pública que con la propia evaluación, en parte explicado porque estaba en marcha la Comisión para la Reforma de las Administraciones públicas (CORA)[3].

En 2017, en un contexto de racionalización del sector público, la AEVAL se disolvió y se creó en su sustitución el Instituto para la Evaluación de Políticas Públicas (IEPP) con rango de subdirección general adscrita a la Secretaría de Estado de Función Pública[4].

---

[3] La CORA, creada en 2012 en el contexto de la Gran Recesión, tenía por objeto eliminar duplicidades y reforzar los mecanismos de cooperación entre Administraciones, avanzar en la simplificación administrativa, identificar las actividades de gestión en las que pudieran optimizarse el uso de los recursos públicos y revisar el número de entes en las Administraciones.

[4] Artículo 13.7 y la disposición adicional novena Real Decreto 769/2017. https://www.boe.es/buscar/pdf/2017/BOE-A-2017-9012-consolidado.pdf

3.1.2. Capacidad evaluadora y cultura de evaluación

Si se compara con otros países, la AEVAL era una Agencia pequeña. Aunque llegó a tener 59 efectivos en 2009, siempre estuvo lejos de su dotación potencial (75 efectivos para ese año) y fue poco a poco perdiendo personal. En 2016, un año antes de su desaparición, contaba con 38 personas (AEVAL, 2018). Además, solo una parte reducida de su personal, menos de la mitad, se dedicaba a la EVPP.

La AEVAL contaba sobre todo con personal funcionario, lo cual pudo ser una limitación dada la falta de experiencia en la Administración pública española en la evaluación de políticas. Además, la Agencia apostó siempre por realizar sus propias evaluaciones, sin contar con medios externos. A las limitaciones de personal, se le unieron las restricciones de acceso a datos a las que se enfrenta tradicionalmente la evaluación de políticas públicas, la llegada de la crisis y otras razones de tipo político.

3.1.3. La actividad evaluadora y la trascendencia de las evaluaciones

Aunque la AEVAL no era la única institución encargada de realizar evaluaciones, sí era sin duda la principal en esos años. La división de evaluación de la AEVAL llevó a cabo alrededor de 41 evaluaciones, dos de ellas para las comunidades autónomas (CC. AA.), y elaboró algunas propuestas metodológicas, programas de formación[5] y cuatro guías[6].

Respecto a sus relaciones, la AEVAL coordinaba la Red Interadministrativa de Calidad de Servicios Públicos, en la que participaban las CC. AA. y la Federación Española de Municipios y Provincias (FEMP) y en cuyo seno había un Comité de Evaluación de Políticas Públicas. Además, participaba en la Red Europea de Administración Pública (EUPAN) con el resto de los países de la Unión Europea. En cuanto a sus vínculos con la sociedad civil, la AEVAL integró en su Consejo Rector a distintos académicos relacionados con las políticas públicas y colaboraba con la Sociedad Española de Evaluación.

Hubo también otros organismos que realizaron sus propias evaluaciones. Un estudio encargado por AEVAL a la Universidad Carlos III sobre la Evaluación de Políticas y Programas Públicos en la Administración General del Estado (2007-2011) identificó 71 evaluaciones «en sentido estricto» o «verdaderas»[7]. Casi la mitad, 34, habían sido llevadas a cabo por la AEVAL, 18 por la Dirección General de Planificación y Evaluación para el Desarrollo en el Ministerio de Asuntos Exteriores y Cooperación, 6 por parte del Ministerio del Medio Ambiente y Medio Rural y

[5] Véase el Eje Estratégico II: la promoción de la cultura de la evaluación (Informe General de Actuación de AEVAL 2016, 2018).

[6] Además de ellos, la división de calidad de los servicios elaboró 24 informes más y varias guías. Esta división realizaba también otras tareas de consultoría administrativa (certificaciones de evaluaciones organizativas o formación, entre otras).

[7] Según Stufflebeam y Shinkfield (1995) las evaluaciones verdaderas son los procesos que, a partir de la investigación sistemática, proporcionan evidencia sólida sobre políticas, programas o planes públicos e incorporan un juicio de valor sobre la base de criterios explícitos o implícitos y orientaciones prácticas en relación con la intervención evaluada.

Marino, 5 por la Fundación Tripartita para la Formación en el Empleo, 3 por el Centro para el Desarrollo Tecnológico Industrial y otras organizaciones (Ruiz, 2012). De entre todas las evaluaciones, solo una fue *ex ante*.

Para acabar el repaso al periodo comprendido entre 2000 y 2017 y poder concretar el grado de institucionalización de la evaluación en España en esos años, cabe mencionar varios análisis sobre institucionalización de la EVPP que incluyeron el caso de España. El estudio de Jacob, Speer & Furubo (2015) establece distintos grados de madurez en relación con la evaluación de políticas en función de la posición de cada país en nueve dimensiones. España, junto a Irlanda, Italia y Nueva Zelanda, se ubica entre los países con un grado de madurez medio. Por su parte, Stockmann, Meyer & Taube (2020b) elaboraron un índice para analizar el grado de institucionalización de los países. Se aplica a las diferentes dimensiones de la institucionalización que identificaron en su marco conceptual anteriormente mencionado. De acuerdo con este, Suiza estaba claramente por encima de los demás países en los tres sistemas que los autores analizan (político, social y profesional), mientras que Rumanía, Portugal, Polonia, Letonia e Irlanda, se situaban por debajo de la media también en los tres. Por su parte, Alemania, Dinamarca y Finlandia consiguieron situarse por encima de la media en dos de los sistemas, mientras que otros países, entre ellos España, Reino Unido, Italia, Bélgica y la República Checa, lo hacen al menos en uno de ellos (Figura 1), por lo que, de acuerdo con ese estudio, se podría concluir que España posee un grado de institucionalización de la evaluación medio-bajo.

FIGURA 1
LA INSTITUCIONALIZACIÓN DE LA EVALUACIÓN DE POLÍTICAS EN PAÍSES SELECCIONADOS

FUENTE: Stockmann, Meyer y Taube (2020b, p. 513). Z-Scores de los índices estandarizados de los sistemas político, social y profesional.

Aunque estos análisis son sin duda interesantes, ninguno de ellos incluye las últimas e importantes iniciativas en materia de EVPP que han tenido lugar en España ni contienen una referencia sistemática a la situación de la evaluación en las CC. AA. En el próximo epígrafe se intenta resolver esta limitación, utilizando el mismo esquema empleado en esta sección para analizar los avances ocurridos en esta materia en España desde 2017 hasta el otoño de 2021.

3.2. El reciente impulso a la evaluación de políticas públicas en la Administración General del Estado (2017 a 2021): la experiencia de la AIReF

3.2.1. Marco organizativo y normativo

La Autoridad Independiente de Responsabilidad Fiscal (AIReF), que en principio no tenía previsto desempeñar la función evaluadora de políticas, ha tenido un protagonismo creciente en los últimos años. La AIReF se crea con la Ley Orgánica de Estabilidad Presupuestaria de 2013. Su origen tiene lugar, al igual que la mayoría de las instituciones fiscales independientes, como respuesta a la crisis de 2008 para el refuerzo de la estabilidad presupuestaria. Su actividad está regulada en su propio Estatuto Orgánico, aprobado por RD 215/2014, de 28 de marzo y su independencia se garantiza por, entre otros mecanismos, su forma de financiación mediante una tasa, su origen y vinculación con Europa, el modo en que se realizan sus encargos con mayor alcance que el de una legislatura, y el modo de selección de su presidente (en la actualidad presidenta), que requiere la mayoría absoluta del Parlamento y se realiza fuera del ciclo político.

Después de alguna experiencia inicial de la AIReF en materia de evaluación por encargo de las CC. AA., el Gobierno central solicitó a la AIReF en 2017 la evaluación de diversos capítulos del gasto público en varios años, fases y etapas. Este encargo supuso el impulso definitivo a la actividad de EVPP para este organismo, aunque no era ajena en instituciones similares en los países de nuestro entorno (Herrero, 2021). Desde ese momento, la evaluación en la AIReF se regula mediante acuerdos del Consejo de Ministros conocidos como *Spending Reviews* (SR) o solicitudes de evaluación del gasto que el Gobierno central hace a la AIReF en políticas sectoriales de competencia exclusiva o compartida. Una vez recibidos los encargos, de periodicidad anual, AIReF elabora para cada uno un plan de acción donde detalla el contexto, objetivos, ejes, metodologías, bases de datos, gobernanza, publicidad, calendario y presupuesto de las evaluaciones. Estos planes de acción tienen que ser ratificados por el Consejo de Ministros. Además de los encargos de la Administración central, la AIReF también realiza evaluaciones para los Gobiernos autonómicos y locales, habiéndose multiplicado exponencialmente sus estudios para estos niveles de gobierno en los últimos años.

Además de la AIReF, en la Administración central existen otros organismos que realizan evaluaciones o realizan algunas tareas vinculadas a la evaluación. En 2020, nuevamente, el RD 307/2020 adscribió el IEPP a la Secretaría de Estado de Política

Territorial y Función Pública, manteniendo su rango de subdirección general. Adicionalmente, algunos centros gestores realizan evaluaciones internamente o con la ayuda de terceros (Instituto de Evaluación de Políticas Públicas, 2021). Estas evaluaciones, en ocasiones, forman parte de planes plurianuales de evaluación. Por poner algunos ejemplos, la cooperación española cuenta con una política de evaluación, un plan bienal de evaluaciones y una metodología de evaluación[8]. También existe un plan anual de evaluación de la formación del Servicio Público de Empleo. La evaluación de este sistema está regulada en la ley 30/2015[9]. En el Ministerio de Trabajo y Economía Social existe un Plan Estratégico Nacional del Fondo Social Europeo[10]. Igualmente, en el caso del Ministerio de Hacienda y Función Pública se dispone de un Plan de Evaluación de los Programas Operativos FEDER[11].

La normativa nacional prevé también varios organismos o dispositivos sectoriales que realizan algún tipo de evaluación, aunque es cierto que no siempre realizan EVPP en sentido estricto. En el ámbito de la educación y la ciencia, pueden mencionarse el Instituto de Evaluación, la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA), la cual evalúa el contenido de las titulaciones universitarias o del perfil del profesorado, o el Sistema Unificado de Evaluación de Investigación, Desarrollo e Innovación. En Sanidad, la Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud también lleva a cabo evaluaciones. Cabe destacar también al Instituto de Estudios Fiscales.

Con todo, el propio Plan de Recuperación, Transformación y Resiliencia (PRTR) presentado por el Gobierno de España en mayo de 2021, con el objetivo de aprovechar los nuevos instrumentos comunitarios de financiación denominados *Next Generation EU,* puestos en marcha en el contexto de la COVID-19, ha considerado que la evaluación necesita un apoyo mayor. El PRTR menciona la EVPP en varios de los denominados componentes, en concreto en el 11 (Modernización de las Administraciones públicas) y en el 29 (Mejora de la eficacia del gasto público) y, en particular, consolida el papel que tendrá AIReF como organismo evaluador reforzando su independencia, profesionalidad, sistematicidad y la relevancia de sus evaluaciones[12].

Además, desde un punto de vista normativo, en la Administración del Estado, en verano de 2021, finalizó el plazo de consulta pública sobre el Anteproyecto de ley de institucionalización de la EVPP en la Administración General del Estado, cuyo

---

[8] http://www.exteriores.gob.es/Portal/es/SalaDePrensa/Multimedia/Publicaciones/Documents/Cooperacion/Evaluacion/Metodologias/Metodologiade%20evaluaci%C3%B3n%20%20de%20la%20Cooperaci%C3%B3n%20Espa%C3%B1ola%20I%20completo.pdf

[9] https://www.sistemanacionalempleo.es/evaluacion_formacion.html

[10] https://www.mites.gob.es/uafse/es/evaluaciones/index.htm En la página se pueden encontrar también las evaluaciones para todas las CC. AA.

[11] https://www.dgfc.sepg.hacienda.gob.es/sitios/dgfc/es-ES/ipr/fcp1420/e/PEF1420/Documents/Plan_de_Evalaucion_Comun_FEDER_v14.pdf

[12] Véase el PRTS https://www.lamoncloa.gob.es/temas/fondos-recuperacion/Documents/30042021-Plan_Recuperacion_%20Transformacion_%20Resiliencia.pdf

objetivo era el de «impulsar la institucionalización y conformar un sistema más robusto de evaluación» por parte del Ministerio de Hacienda y Función Pública[13].

3.2.2. Capacidad evaluadora y cultura de evaluación

El presupuesto público dedicado a evaluación es creciente, pero las Administraciones se siguen viendo limitadas por otros aspectos que condicionan su capacidad de evaluar. En cuanto a la disponibilidad de efectivos para llevar a cabo las tareas de evaluación, en el caso del IEPP, su Relación de Puestos de Trabajo (RPT) incluye once efectivos. Por su parte, en AIReF, el personal dedicado a la evaluación es escaso, sin sobrepasar en la actualidad la decena de efectivos (con perfiles de economía, matemáticas, estadística y ciencia política). Aunque AIReF asume en solitario una parte de las evaluaciones que realiza, para otras, cuenta con medios externos. Pero, incluso en este último caso, el volumen de trabajo es grande debido a que la AIReF diseña, dirige y coordina sus evaluaciones y se involucra intensamente en el proceso de evaluación mediante la configuración de equipos mixtos con terceros actores. Esta práctica permite al propio organismo aprender de cada evaluación, además de contribuir a generar un mercado de evaluación (en universidades y el sector privado), lo cual es importante para que se pueda hablar de institucionalización de la evaluación en un país, tal y como se ha señalado anteriormente.

Por su parte, existen varias iniciativas sobre evaluación en la sociedad civil, aunque aún son excepcionales en España. En la academia, algunos programas de grado y postgrado incluyen asignaturas de evaluación de políticas, por ejemplo, las carreras de Economía o Ciencias Políticas de algunas facultades, siendo también común en másteres en gestión y Administración pública. Por lo que se refiere a otras iniciativas no estrictamente académicas, cabe destacar nuevas redes que integran a expertos y académicos con capacidades cuantitativas y cualitativas para realizar evaluaciones[14]. Además, existen organizaciones con larga tradición como la Sociedad Española de Evaluación de Políticas y la Asociación Ibérica de Profesionales por la Evaluación (APROEVAL). Estas asociaciones y sus miembros participan activamente en reuniones científicas nacionales e interacciones dedicadas a la difusión de la evaluación de políticas o donde la evaluación ocupa algún tipo de espacio. En la propia Administración, el Grupo de Referencia en Evaluación de la Administración General del Estado se reúne mensualmente. Quizá en un futuro cercano se podrá hablar de un mercado interno de evaluación, en el que distintos organismos de las Administraciones se especialicen en distintos tipos de evaluación sectorial.

En la Administración central, el PRTR (p. 172) ha puesto de manifiesto el propósito de apoyar la función evaluadora de la AIReF dada la creciente demanda

[13] Este Anteproyecto de ley puede consultarse aquí: https://www.mptfp.gob.es/portal/ministerio/participacion_proyectos/consulta_previa/proyectos/2021/2021-05-17.html

[14] Por ejemplo, la Red por las Buenas Prácticas en Evaluación (RedEV), integrada por cuatro organizaciones (Afi, Iseak, Ivie y Ksnet).

mediante «el refuerzo de la capacidad del evaluador a través de la creación de una división permanente de evaluación en el seno de la AIReF para dar continuidad y permanencia a los ejercicios de *Spending Review*». La División de Evaluación del Gasto Público se ha creado en septiembre de 2021, consolidando así, en la estructura de la AIReF, una función que ya venía realizándose.

En relación también con la capacidad, uno de los asuntos que más se han discutido en los últimos años ha sido la disponibilidad de datos para la realización de las evaluaciones. En este sentido, la AIReF propuso el diseño de una estrategia nacional para avanzar en la disponibilidad de datos administrativos convenientemente anonimizados, para ponerlos a disposición de las Administraciones y de la academia en general[15], que fue compartida también por la comunidad científica e investigadora (Almunia & Rey-Biel, 2020).

3.2.3. La actividad evaluadora y la trascendencia de las evaluaciones: nuevos avances a raíz del PRTR

La ausencia de un observatorio que recoja evaluaciones y recomendaciones, convierte en muy difícil la tarea de realizar un seguimiento exhaustivo de la actividad evaluadora de decenas de estos organismos y de la utilización de estos análisis[16].

Entre 2017 y 2020, la actividad del IEPP ha sido limitada, realizando una evaluación y cuatro guías[17]. En cuanto a la AIReF, en 2017, el Gobierno de España encargó a este organismo la revisión integral del gasto público (el ya mencionado aquí SR). Este ejercicio, con una duración prevista de tres años, se ha concretado en tres encargos (Cuadro 1). En el marco del SR, las evaluaciones solicitadas por la Administración General del Estado se han referido a importantes capítulos de gasto y suponen algo más de 75.330 millones de euros (alrededor de un 15 % del gasto público total). Entre los sectores analizados se encuentran subvenciones, sanidad, beneficios fiscales, empleo, becas, I+D, industria, infraestructuras, correos y telégrafos, o lucha contra la pobreza. El catálogo de políticas a evaluar para la Administración central se ha ampliado con la tercera fase del SR, que incluye las líneas de préstamos, ayudas e instrumentos financieros al tejido empresarial por parte de las entidades públicas y la gestión de residuos, así como un estudio de evaluación relativo a la cooperación financiera internacional y, en particular, del Fondo para la Promoción del Desarrollo (FONPRODE), encargado también a la AIReF por el Consejo de Ministros.

---

[15] Véase la *Opinión para una estrategia de acceso a datos administrativos*. https://www.airef.es/es/opiniones/estrategia-acceso-datos-administrativos/

[16] En varias de las direcciones web de los organismos mencionados en el epígrafe anterior pueden encontrarse los informes de evaluación, en algunos casos uno al año.

[17] Véanse los informes del IEPP en https://www.mptfp.gob.es/portal/funcionpublica/evaluacion-politicas-publicas/Informes-de-Evaluacion.html

A lo largo de este tiempo, los Gobiernos autonómicos y un ayuntamiento han solicitado a la AIReF diferentes evaluaciones en diversos sectores de política de gran relevancia como la sanidad, la educación, o el empleo.

En total, desde 2017 y hasta el momento, la AIReF ha realizado tanto para el Gobierno central como para nueve comunidades autónomas, quince evaluaciones. Están en curso otras once y varias más se encuentran en el proceso de solicitud (Cuadro 1).

CUADRO 1

EVALUACIONES COMPLETADAS POR LA AIREF O EN FASE DE DESARROLLO POR ENCARGO DEL GOBIERNO CENTRAL Y DE LAS COMUNIDADES AUTÓNOMAS

ENCARGO PLURIANUAL DEL GOBIERNO CENTRAL: SPENDING REVIEW

Fase 1. Spending Review (2018-2019)

1	Evaluación de estrategia y procedimiento en la concesión de subvenciones
2	Medicamentos dispensados a través de receta médica
3	Políticas activas de empleo
4	Becas de educación universitaria
5	Promoción de talento y empleabilidad en I+D+i
6	Fortalecimiento de la competitividad empresarial
7	Sociedad Estatal Correos yTelégrafos

Fase 2. Spending Review (2019-2020)

1	Beneficios fiscales
2	Gasto hospitalario del SNS: farmacia e inversión en bienes de equipo
3	Incentivos a la contratación y al trabajo autónomo
4	Infraestructuras de transporte

Fase 3. Spending Review (2021-)

1	Líneas de préstamos, ayudas e instrumentos financieros al tejido empresarial por entidades públicas
2	Gestión de los residuos urbanos

OTROS ENCARGOS DEL GOBIERNO CENTRAL

Estudio sobre **Renta Mínima, Ingreso Mínimo Vital, Fondo para la Promoción del Desarrollo FONPRODE**

ENCARGOS DE EVALUACIONES DE LOS GOBIERNOS AUTONÓMICOS (9 CC. AA.)

Sanidad: farmacia hospitalaria, extrahospitalaria, sociosanitaria, inversión alta tecnología, sistemas información, recursos humanos, determinantes gasto
Políticas activas de empleo: orientación, incentivos, formación, inserción, itinerarios, etc.
Educación: universidades y no universitaria, gobernanza, actividad productiva, financiación, recursos humanos
Política industrial
Institucionalización de la evaluación: diagnóstico y propuestas
Publicidad institucional, gobernanza y planificación, transferencias de valor a los medios, transparencia
Estudio sobre empresas y fundaciones públicas autonómicas
Concesiones para la construcción y explotación de infraestructuras

FUENTE: Elaboración propia.

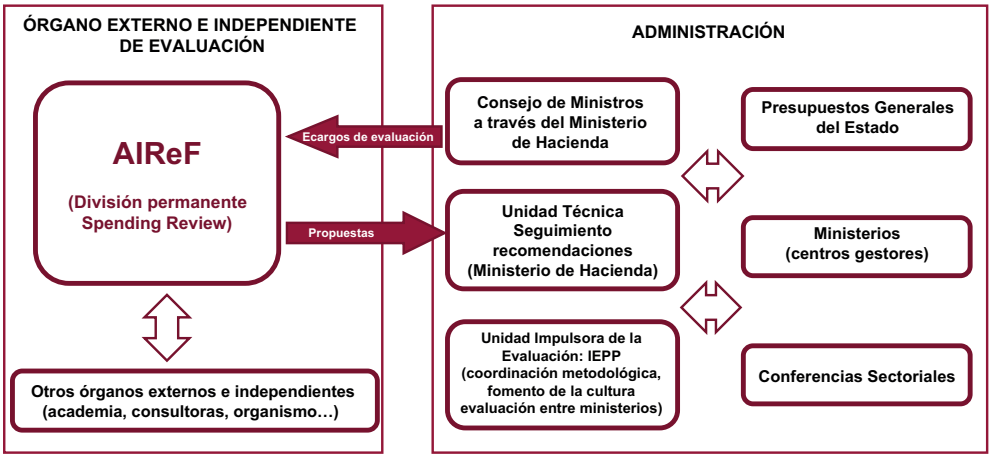
Además, en 2021, el PRTR ha incluido un hito de gran importancia para la consolidación del proceso de evaluación del gasto en España planteando un nuevo proceso de Spending Review para un período de cinco años (2022-2026). Este periodo pone de manifiesto la intención de separar el proceso de evaluación de las legislaturas.

Tan importante como realizar las evaluaciones es asegurar su utilización para lo que, incluso en los sectores donde la EVPP cuenta con más tradición, como la cooperación al desarrollo, este sigue siendo un reto pendiente (Gudiño, 2015). El propio PRTR (p. 133) prevé algunas reformas tomando como referencia los resultados de las evaluaciones de AIReF en el ámbito de las políticas activas de empleo, la simplificación y racionalización de los incentivos a la contratación o los beneficios fiscales. Otras evaluaciones de AIReF han contribuido a la configuración de distintas medidas de política pública en industria, fiscalidad, sanidad o universidades, o en relación con el Ingreso Mínimo Vital.

También vinculado a garantizar el uso de las evaluaciones, otro importante hito recogido en el PRTR es la previsión de creación de una unidad técnica en el Ministerio de Hacienda para el seguimiento de las propuestas de la AIReF. Se avanzará así en la implementación del principio de «cumplir o explicar», según el cual las Administraciones deben al menos estudiar las propuestas de la AIReF y razonar su no-aplicación. También debe destacarse la propuesta que incluye el PRTR, consistente en que las evaluaciones sean llevadas a las conferencias sectoriales donde se reúnen cada ministerio y las consejerías del ramo. Además, se prevé que, antes del 31 de marzo de cada año, el Ministerio de Hacienda remita al Consejo de Ministros un informe con el estado de las recomendaciones. La Figura 2 resume la arquitectura institucional que, de aplicarse las medidas contenidas el PRTR, involucrará a la AIReF con otras instituciones del Estado y, por tanto, supondrá un paso importante para la institucionalización de la evaluación.

FIGURA 2

LA ARQUITECTURA INSTITUCIONAL EN MATERIA DE EVALUACIÓN PREVISTA EN EL PRTR (2021)



FUENTE: Elaboración propia.

Finalmente, de cara a garantizar la trascendencia de las evaluaciones, se ha aprobado la Orden del Ministerio de Hacienda en la que se establecen las normas para la elaboración de los Presupuestos Generales del Estado para 2022. En ella se determina que los resultados de los procesos de evaluación del gasto público serán tenidos en cuenta a partir de ahora en la discusión presupuestaria y, para ello, la Dirección General de Presupuestos podrá recabar la información que requiera a los centros gestores de los distintos ministerios.

3.3. La evaluación de políticas públicas en las comunidades autónomas (2005 a 2021)

3.3.1. Marco organizativo y normativo

Las CC. AA. también han reforzado sus marcos legales y organizativos para evaluar las políticas. Aunque existen una gran variedad de situaciones que se expresan en la distinta plasmación normativa y que, en la mayoría de los casos, se circunscribe a reforzar la evaluación para el seguimiento de las políticas dentro de los centros gestores de ámbito sectorial.

Navarra cuenta con la Ley Foral 21/2005 de Evaluación de Políticas Públicas y Calidad de los Servicios. Esta ley, promulgada antes de la creación de la AEVAL, contiene una regulación detallada que declara la voluntad de constituir y desarrollar un sistema público de evaluación y establece que cada organismo podrá acordar autónomamente la evaluación de sus políticas.

Sin disponer de una ley, otras CC. AA. dedican parte del articulado de algunas de sus normas a la EVPP. En Aragón, la Ley 5/2017 de Integridad y Ética Públicas, completa las referencias a la evaluación que ya hacía su Ley 5/2013 de Calidad de los Servicios, estableciendo quién será el organismo evaluador (Agencia de Integridad y Ética Públicas), los principios de evaluación y la rendición de cuentas debida a las Cortes de Aragón. En Cataluña, la Ley 5/2020 de medidas fiscales, financieras, administrativas y del sector público, añade un capítulo al texto refundido de su Ley de Finanzas Públicas en el que define la EVPP, los criterios de evaluación y sus objetivos, así como el proceso de evaluación en el que participarán el departamento competente en materia de finanzas públicas y los centros directivos, que deberán realizar planes de evaluación plurianuales de las políticas que hayan diseñado. En Galicia, el Decreto 97/2020 regula determinados instrumentos para la gestión, evaluación, supervisión y mejora de la calidad en el sector público autonómico y regula las modalidades de lo que denomina evaluación de la calidad de las políticas públicas. Castilla y León también cuenta con varias normas pero, como demostró la reciente evaluación de AIReF sobre este tema en la comunidad autónoma, no están implementadas hasta el momento (AIReF, 2021).

La Junta de Andalucía trabaja en la Ley de Organización y Funcionamiento del Sistema de Evaluación de Políticas Públicas para establecer el marco organizativo

necesario para la evaluación y regular su funcionamiento e interrelación entre las fases del ciclo de las políticas, así como la participación ciudadana en cada etapa. El Plan Estratégico de Gobernanza e Innovación Pública 2020 del País Vasco prevé la implantación del Sistema de Evaluación de las Intervenciones Públicas.

Estas, y gran parte del resto de las CC. AA., incluyen referencias a la EVPP en las leyes de transparencia, de participación o Gobierno Abierto que han sido aprobadas desde 2011. Sin embargo, en la mayoría de las ocasiones se trata de menciones generales como, por ejemplo, al derecho de los ciudadanos a participar en las evaluaciones de las políticas. Por otro lado, en todas las CC. AA. existe regulación sectorial que realiza previsiones sobre la EVPP, por ejemplo, es habitual en la normativa de políticas activas de empleo, de protección ambiental, en las políticas sanitarias, educativas, de género, de servicios sociales o en las políticas de transparencia.

Por lo que se refiere a la estructura organizativa de la evaluación en las CC. AA., la situación también es diversa. Existen pocas agencias generalistas, siendo el caso de Iválua en Cataluña el más conocido. Este consorcio creado en 2006 está formado por la Generalitat, la Diputación de Barcelona y la Universidad Pompeu Fabra. En 2019, Navarra estableció que el Observatorio de la Realidad Social ejercerá la planificación y la evaluación de las políticas de servicios sociales y empleo, así como el análisis de los procesos de transformación social y de los problemas sociales.

En Galicia, existe una Comisión Interdepartamental de Información y Evaluación que determinará los planes que deben ser objeto de evaluación externa y a la que se remitirán las evaluaciones. Se ha creado recientemente el Instituto de Economía Pública de Galicia, al que se le atribuirán algunas funciones de evaluación. En Andalucía, el Instituto Andaluz de Administración Pública proporciona servicios vinculados a la evaluación, por ejemplo, análisis de la evaluabilidad de los planes estratégicos, evaluaciones *ex ante,* labores de difusión, una red corporativa de evaluadores y varias guías. Fruto de la evaluación realizada por AIReF, Castilla y León ha anunciado la intención de crear una Agencia Autonómica de Evaluación Independiente de las Políticas Públicas y la Calidad de los Servicios adscrita a su Consejo de Cuentas.

Numerosas CC. AA. disponen en sus organigramas de unidades responsables de la evaluación de políticas de diferentes sectores. Por ejemplo, en el caso de Extremadura, a la Dirección General de Planificación y Evaluación de las Políticas Activas del Servicio Extremeño de Políticas de Empleo le corresponde la planificación y evaluación de las políticas activas de empleo. En la Comunidad de Madrid puede encontrase un Área de Evaluación de Políticas Activas y Estrategia de Empleo en la Consejería de Empleo, así como varias subdirecciones generales de evaluación en consejerías sectoriales, como en educación o en políticas sociales. En algunas CC. AA. existen agencias de evaluación de la educación, como la Agència Valenciana d'Avaluació i Prospectiva o la Agència de Gestió d'Ajusts Universitaris i de Recerca en Cataluña. En otras CC. AA., como el Principado de Asturias y Cantabria, las evaluaciones disponibles están vinculadas a servicios o programas concretos, como los servicios sociales o los programas de género. En La Rioja, la Consejería de Hacienda incluye en su organigrama un Servicio de Evaluación de Políticas.

En 2018, Aragón constituyó la Comisión para la Evaluación de las Políticas Públicas y la Calidad de los Servicios Públicos. En las Islas Baleares se prevé la creación de una Oficina de Evaluación Pública, que está dando los primeros pasos. En el caso de Canarias, dentro de la Estrategia para la Modernización y Mejora de los Servicios Públicos (PEMAC 2012-2014), se desarrollaron diferentes acciones formativas.

En Castilla-La Mancha se ha anunciado la creación de un nuevo órgano propio de fiscalización externa de las finanzas públicas (la Sindicatura de Cuentas se suprimió en 2014), que tiene previsto realizar tareas de fiscalización, analizando la utilización de recursos y «atendiendo al grado de cumplimiento de los objetivos programados», según reza el Anteproyecto de Ley de la Cámara de Cuentas de Castilla-La Mancha.

3.3.2. Capacidad evaluadora y cultura de evaluación

Tres análisis recientes de CC. AA. específicas llegan a conclusiones similares, que probablemente podrían ser compartidas por la mayoría de ellas. Los tres trabajos prueban que en las CC. AA. existen conocimientos y concienciación sobre la utilidad de la evaluación, pero falta reforzar la formación de los empleados públicos, los sistemas de información y el uso de los datos.

En Murcia, y dentro del marco de la Estrategia Regional de Gobernanza Pública, se constituyó un grupo de trabajo sobre EVPP, que en 2020 presentó un diagnóstico de la EVPP en la región con resultados interesantes sobre la capacidad de evaluación (carencia de recursos, ausencia de metodología y estructura organizativa adecuada, falta de interés, entre otros) y la necesidad de reforzarla[18].

En un estudio sobre la EVPP en la Comunidad Valenciana, se identifican numerosas prácticas de evaluación en la Generalitat e incluso diferentes fases del proceso de institucionalización según la consejería o el organismo (Barberà, Doria, Ntutumu-Sanchis & Sanchis, 2020). Destaca el arraigo de la evaluación en los casos de la Agencia Valenciana d'Avaluació i Prospectiva o en el área de cooperación al desarrollo. Además, en la propia Consellería de Presidencia se encuentra la Dirección General de Análisis de Políticas Públicas. Existen numerosas guías, pero no un enfoque único de evaluación para toda la Administración autonómica. Se detectan también importantes necesidades de mejora, especialmente, en la capacidad evaluadora.

AIReF (2021) ha estudiado recientemente el caso de Castilla y León, concluyendo que en esta comunidad podría implementarse una estructura para la EVPP aprovechando la experiencia de los organismos ya existentes y con escasas modificaciones legales. Sin embargo, también concluyó que es necesario reforzar elementos clave de la capacidad, en particular los sistemas de información y el uso de datos y la formación de los empleados públicos.

---

[18] https://transparencia.carm.es/wres/transparencia/doc/Organos-colegiados/Grupo_de_trabajo_Eval_pol_pub/2020_12_03/Mapa_Diagnostico_Calidad_Evaluacion_y_Planificaci%C3%B3n_CARM.pdf

En esta comunidad, un 66 % de los responsables públicos entrevistados indica que los planes y programas incorporan indicadores de actividad o de implementación, pero solo la mitad de los entrevistados afirman que disponen de indicadores de impacto. El tratamiento informático de los datos también es limitado (menos del 40 % señalan que esto se haga muy a menudo o a menudo). Solo un 25 % afirma que explota los datos muy a menudo o a menudo y el cruce entre datos en diferentes consejerías de la propia Junta es infrecuente (solo un 12 % reconoce que esto se hace muy a menudo o a menudo). No existe una plataforma de datos donde se posibilite el acceso fácil, ágil y seguro en términos de confidencialidad por parte de la propia Administración autonómica a la información de las distintas consejerías.

En cuanto a las necesidades formativas vinculadas a las tareas de evaluación, al menos un 40 % de los responsables públicos asegura que sus equipos han tenido que realizar muy a menudo o a menudo tareas como diseñar indicadores de seguimiento e impacto, realizar informes de seguimiento de la implementación y el impacto de los programas o diseñar algún tipo de evaluación. Pero, al mismo tiempo, reconocen que, para muchas de estas tareas que los empleados públicos ya realizan, no cuentan con la formación adecuada. La Figura 3 muestra que, en general, a medida que las tareas que toda evaluación implica se van sofisticando, los responsables públicos creen que su equipo requiere un apoyo formativo mayor. Por ejemplo, alrededor de un 80 % del personal necesitaría formación para poder establecer indicadores de implementación e impacto de los programas. El uso de metodología cualitativa y, especialmente, de metodología cuantitativa y experimental supone otra necesidad para los trabajadores respecto a la cual alrededor del 90 % de los responsables públicos identifican que se requiere mayor formación.

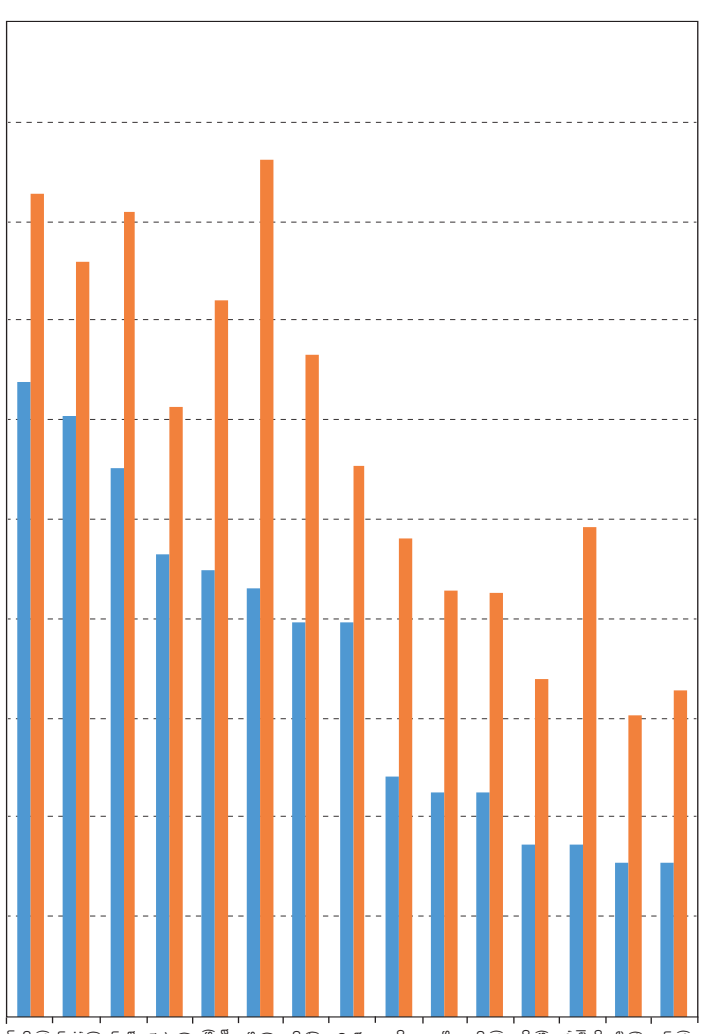
A pesar de que el 98 % de los responsables públicos entrevistados creen que la evaluación es una fuente de información para el diseño de políticas, un elevado 85 % afirma utilizar muy a menudo o a menudo la intuición, y menos de un 35 % utiliza alguna fuente de conocimiento externo al programa.

Otros objetivos de la evaluación para los entrevistados son: rendir cuentas a la ciudadanía (95 %), el aprendizaje organizativo (90 %) y la eficiencia en el gasto (86 %), ayudar para tomar decisiones políticamente delicadas (72 %) y entender las necesidades de la ciudadanía (71 %). Además, creen que debe dársele más importancia a la EVPP en sus consejerías. Utilizando una escala de 0 a 10, consideran que la importancia que tiene la evaluación supera por poco el aprobado (5,4) y creen que debería alcanzar al menos el 8,2.

Para acabar este repaso de las Administraciones autonómicas, cabe destacar la participación de las CC. AA. en el IV Plan de Gobierno Abierto de España 2020-2024, elaborado por la Administración General del Estado, las propias Administraciones regionales y la FEMP. Este plan se compromete a avanzar en la gestión de datos en abierto, lo que es importante para la EVPP. Además, algunas instituciones de control externo autonómicas han reforzado la formación de su personal y han realizado, junto a sus habituales auditorías de gestión u operativas, algunas evaluaciones.

FIGURA 3

FRECUENCIA CON LA QUE EN LOS CENTROS DIRECTIVOS DE UNA ADMINISTRACIÓN AUTONÓMICA SE REALIZAN TAREAS VINCULADAS CON LA EVALUACIÓN Y NECESIDADES DE FORMACIÓN RELACIONADAS CON ELLAS (2021)



FUENTE: AIReF (2021). Cuestionario a los coordinadores de la Junta de Castilla y León. Pregunta: «Indique en qué medida su unidad o centro directivo ha realizado o ha encargado que se realicen las siguientes actividades en los últimos cinco años y pensando en que su centro directivo o unidad tuviera que hacerse cargo de alguna de las siguientes actividades, dígame por favor, si su equipo necesita formación complementaria para hacer adecuadamente estas tareas».

3.3.3. La actividad evaluadora y la trascendencia de las evaluaciones

Como ocurría en el caso de la Administración General del Estado, es difícil identificar las evaluaciones que se llevan cabo en las CC. AA., salvo cuando cuentan con una institución de evaluación. En el caso de Cataluña, Ivàlua recoge en su página web sus informes de evaluación y lo que llama informes de revisión de la evidencia[19]. El Observatorio de la Realidad Social de Políticas de Navarra tiene también disponibles algunos estudios.

Aunque las CC. AA. no cuentan con instituciones de evaluación propias, en ocasiones encargan a terceros estos ejercicios. Hay que recordar que ocho CC. AA. (Cantabria, Cataluña, Andalucía, Castilla y León, Aragón, Extremadura, la Comunidad Valenciana y Navarra) han realizado varios encargos de evaluación a la AIReF, cada una de ellas en diversos sectores de política pública.

Asimismo, la mayoría de las CC. AA. disponen de guías de evaluación generales (sobre diseño de la evaluación o distintos tipos de evaluación) o sectoriales (por ejemplo, de evaluación de políticas sociales o de género, entre otros). Ivàlua cuenta con una gran variedad de ellas. El caso más reciente es el de Murcia que, en 2021, ha presentado la Guía de Evaluación de Planes, Programas, Estrategias y Servicios Públicos, elaborada en colaboración con la Universidad de Murcia.

En cuanto a la trascendencia de la evaluación, puede ser de interés la reciente innovación del caso de Cataluña, quien en su normativa de 2020 (mencionada anteriormente) se refiere de manera específica a lo que denomina revisiones de gasto que tienen «por objeto detectar ganancias de eficiencia y eficacia en la gestión de los recursos públicos para maximizar el impacto social del gasto público». En este caso, el departamento competente en materia de economía y finanzas realizará este proceso de revisión del gasto, en colaboración con el resto de los departamentos y entidades del sector público de la Generalitat.

4. Discusión y conclusiones

En este artículo se ha repasado en qué medida se encuentran en España los elementos que están presentes en los sistemas de evaluación de otros países que han avanzado más en el proceso de institucionalización. Se ha revisado hasta qué punto existe un marco normativo y organizativo que ampare la evaluación, se dispone de capacidad evaluadora y pueden encontrarse entre los responsables públicos actitudes y valores afines a la evaluación (una cultura de evaluación), y si se practican y utilizan las evaluaciones. En este epígrafe, se recapitulan los hallazgos y se discute si estos elementos logran garantizar los principios de independencia, investigación sistemática, profesionalidad y relevancia, que la literatura sobre institucionalización

[19] https://ivalua.cat/es/node/31

de evaluación considera esenciales para el buen funcionamiento de los sistemas de EVPP.

La primera conclusión es que los mencionados elementos están presentes de una u otra manera en España. La estructura institucional de evaluación está compuesta por la AIReF, configurándose como el principal agente evaluador en los últimos años, el IEPP y otros organismos públicos y agentes privados que practican la evaluación de políticas de una forma más o menos sistemática. También algunas CC. AA. cuentan con entes dedicados a la evaluación y/o encargan evaluaciones a otras entidades como AIReF. En relación con el marco normativo, no existe una ley de evaluación, rasgo que se comparte con otros países, incluso con algunos en los que la EVPP es una actividad con cierta tradición, como Canadá o Reino Unido (OCDE, 2021), aunque se ha abierto una consulta pública sobre el tema. En el caso de AIReF, las reglas de sus evaluaciones se establecen en los acuerdos (planes de acción) que se firman en el contexto de cada encargo que AIReF recibe por parte de los Gobiernos. Aunque no hay normativa que asegure las evaluaciones, como hemos visto, la AIReF ya ha recibido tres encargos de evaluación del Gobierno central y ocho de las CC. AA., cada uno de los cuales implica varias evaluaciones, y existe el compromiso de continuar realizando encargos hasta el 2026 (SR 2022-2026, Componente 29 del PRTR). Además, en el nivel central de gobierno también existen algunos planes sectoriales de evaluación que parecen haberse consolidado.

Cierta formalización legal es necesaria, en particular para intentar garantizar la continuidad y calendarizar mejor los encargos de evaluación, así como para aclarar el papel de los organismos implicados, por ejemplo, a la hora de decidir quién puede proponer qué políticas deben ser evaluadas. Sin embargo, como demostró la experiencia de la AEVAL, el despliegue normativo no es suficiente. Además, una reciente evaluación de la AIReF sobre la institucionalización de la EVPP en Castilla y León ha encontrado numerosas previsiones normativas que permiten e incluso obligan a poner en marcha un sistema de evaluación. Sin embargo, en muchas ocasiones las normas no se han implementado y los sistemas no están funcionando.

En España, la evaluación es una práctica relativamente reciente impulsada en nuestro caso por la pertenencia europea, tanto en sus primeros pasos como en sus más recientes avances, en particular en el caso de la AIReF (también de origen y fuerte vinculación europeísta). Como también hemos visto en el repaso del caso español, la consolidación del sistema requiere, sin duda, del apoyo político y, en particular, este apoyo debe, paradójicamente, garantizar que los evaluadores gozarán de independencia respecto de aquellos que tienen que garantizársela.

En ocasiones se ha discutido la conveniencia de vincular las agencias evaluadoras al Parlamento para garantizar su independencia. Sin embargo, esta es una opción que también presenta inconvenientes (como la lejanía respecto de la Administración evaluada, los escasos recursos con los que cuentan algunos parlamentos y la escasa formación en materia de análisis de políticas públicas de gran parte de su *staff*). La supuesta ventaja que tendría la vinculación al Parlamento puede garantizarse y, de hecho, se garantiza en diferentes países de otros modos. En el caso de la AIReF,

el sistema de elección de su presidenta por mayoría absoluta del Parlamento, su origen y vinculación con Europa, su sistema de financiación y el encargo de los SR separados del ciclo político contribuyen a garantizar el principio de independencia. Aunque AIReF no puede, por el momento, decidir sobre las evaluaciones que desea hacer, en la práctica y en la actualidad, hay que reconocer el gran interés que los Gobiernos están mostrando en evaluar sus programas y políticas. Por otro lado, la experiencia demuestra que la petición voluntaria de los Gobiernos para que sus políticas sean evaluadas facilita la consecución de datos, la constitución de equipos mixtos compuestos por evaluados y evaluadores y el aprendizaje, que no solo se produce alrededor del resultado de las evaluaciones, sino también a lo largo del proceso. Igualmente, es de esperar que cuando las Administraciones solicitan evaluaciones sean más proclives a aceptar las propuestas de cambio o reforma.

Otro elemento imprescindible para la institucionalización y que también depende en gran medida de la voluntad política es la difícil construcción de la capacidad evaluadora que garantice el principio de profesionalidad e investigación sistemática. La EVPP requiere de una política de evaluación que tenga continuidad en el tiempo (frente a intervenciones esporádicas), es decir, de un presupuesto propio y herramientas como tecnologías, datos y sistemas de información, que van mejorándose progresivamente. Como ya se ha explicado, tanto en el nivel central de gobierno como en las Administraciones autonómicas este asunto sigue presentando un amplio margen de mejora.

En particular, tanto si las Administraciones hacen evaluación como si pretenden solo gestionar encargos que hagan a terceros, es necesario un personal que disponga de habilidades y de un elevado nivel de cualificación interdisciplinar, que requiere, además, estar en continuo aprendizaje debido a que las tecnologías, las metodologías y los enfoques cambian continuamente (Vedung, 2010). El análisis de una comunidad autónoma ha detectado importantes necesidades de formación básica y especializada en evaluación, que probablemente serán similares a las de muchas otras CC. AA. y ministerios.

Respecto a la cultura evaluadora, los responsables públicos entienden la utilidad de la evaluación, pero es necesario que los gestores públicos comprueben las ventajas de contar con la evidencia en las decisiones de política pública (y el menor protagonismo que tiene que tener la intuición) y se entienda que la evaluación implica estimar el impacto de las intervenciones públicas en los problemas sociales de forma objetiva, y si existe una forma más eficaz y eficiente de resolver estos, y no solo en evaluar la legalidad del gasto o si las políticas se están implementando como estaba previsto.

Por último, disponer de una herramienta que reuniese todos los ejercicios de evaluación y el alcance y trayectoria de sus recomendaciones permitiría también gestionar el conocimiento producido, y que las Administraciones y otros actores pudieran intercambiar una valiosa información para el diseño de programas públicos. La AIReF tiene ya programada la puesta en marcha de un observatorio que recoja y sistematice el resultado de sus propias evaluaciones y el seguimiento de las propuestas que realiza a cada Administración. El reciente PRTR y el propio Ministerio de

Hacienda han aprobado algunas medidas para asegurar el uso de las evaluaciones de AIReF en el proceso presupuestario. Sin embargo, asegurar el principio de cumplir o explicar es el gran reto pendiente. Como ya se ha explicado, de acuerdo con este principio los centros directivos cuyas políticas se han evaluado tienen o bien que incorporar los resultados de la evaluación al rediseño de sus políticas o bien, si no los incorporan, deben explicar por qué no lo hacen. Junto a las medidas que se acaban de mencionar, para garantizar este principio es necesario avanzar en otras como la publicación incondicional de los informes de evaluación y su debida difusión en la sociedad civil y, en particular, entre los parlamentarios.

Para acabar, como hemos experimentado en España, la institucionalización de la EVPP es un proceso a menudo tortuoso, con frecuentes avances y retrocesos. En los dos análisis internacionales que se han presentado en este artículo, la institucionalización del sistema de evaluación español se ubicaba en una posición intermedia. Ninguno de los dos análisis incluye los recientes avances en materia de evaluación que España ha experimentado en los últimos años, en particular el surgimiento y refuerzo de la AIReF en el terreno de la evaluación de políticas y los avances en las CC. AA.

Con todo, a pesar de que al incluirse estos recientes avances es posible que la posición de España mejore en los *rankings* internacionales sobre la institucionalización de evaluación de políticas, precisamente por lo difícil que en nuestro país está resultando la consolidación del sistema de evaluación, se hace necesario no caer en la complacencia y continuar trabajando para seguir avanzando con un enfoque realista, pragmático y posibilista que dé respuesta a los retos que se han mencionado en este artículo.

Referencias bibliográficas

AEVAL, Agencia Estatal de Evaluación de las Políticas Públicas y Calidad de los Servicios. (2018). *Informe General de Actuación de AEVAL 2016*. Ministerio de Hacienda. http://www.aeval.es/export/sites/aeval/comun/pdf/agencia/Informe_general_actividad_2016_AEVAL.pdf

AIReF, Autoridad Independiente de Responsabilidad Fiscal. (2021). *La institucionalización de la evaluación de Políticas Públicas en Castilla y León: Situación actual y propuestas*. https://www.airef.es/wp-content/uploads/2021/10/ESTUDIOS/PDF-WEB-accesible-CASTILLA-Y-LEON-1.pdf

Almunia, M., & Rey-Biel, P. (2020). *Por un cambio de cultura en la gestión de los datos en España: Una propuesta de reforma*. EsadeECPol, Center for Economic Policy & Political Economy, Insight #17.

Banco Mundial. (2019). *World Bank Group Evaluation Principles*. https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf

Barberà, O., Doria, E., Ntutumu-Sanchis, F., & Sanchis, P. (2020). *La institucionalización de la Evaluación de Políticas Públicas. La Comunitat Valenciana en perspectiva comparada*. Universidat de València.

Bustelo, M. (2020). Spain. In R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 303-328). Palgrave Macmillan.

CEPAL. (2021). *Evaluación de programas públicos*. Serie Gestión Pública 87. https://repositorio.cepal.org/bitstream/handle/11362/46795/1/S2100039_es.pdf

De la Rica, S. (2020, 17 de octubre). La evaluación de las políticas públicas en el siglo XXI. *El País*. https://elpais.com/economia/2020-10-17/la-evaluacion-de-las-politicas-publicas-en-el-siglo-xxi.html

Del Pino, E., & Colino, C. (2021). ¿Qué sabemos sobre cómo reformar la administración?: contenidos, capacidad y trayectorias. In E. Del Pino, & J. Subirats (Eds.), *Repensando la Administración ante los Nuevos Riesgos Sociales y Globales* (pp. 69-100). INAP.

De Peuter, B., & Pattyn, V. (2009). *Evaluation capacity: enabler or exponent of evaluation culture?* In A., Fouquet, & L., Méasson (Eds.), *L'évaluation des politiques publiques en Europe: Cultures et Futurs // Policy and programme evaluation in Europe: Cultures and Prospects* (pp. 133-142). Editions L'Harmattan.

Eurosocial. (2015). *Manual de buenas prácticas en la sistematización de la evaluación de políticas públicas. Experiencias de América latina*. Documento de trabajo n.º 29.

FEDEA, Fundación de Estudios de Economía Aplicada. (2021). *La evaluación de políticas públicas en España: antecedentes, situación actual y propuestas para una reforma*. FEDEA Policy Papers 2021/09. https://documentos.fedea.net/pubs/fpp/2021/10/FPP2021-09.pdf?utm_source=wordpress&utm_medium=portada&utm_campaign=estudio

Feinstein, O., & Zapico-Goñi, E. (2010). *Evaluation of Government Performance and Public Policies in Spain* (ECD Working Papers Series No. 22). The World Bank. Independent Evaluation Group.

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

Garde-Roca, J. A. (2006). La institucionalización de la evaluación de las políticas públicas en España. *Auditoria Pública,* 39, 17-26.

Garde-Roca, J. A. (2017, 5 de Agosto). AEVAL y su muerte anunciada. *Cinco Días*. *El País*. https://cincodias.elpais.com/cincodias/2017/08/04/midinero/1501857354_174399.html

Gobierno de España. (2021a). *Plan de Recuperación, Transformación y Resiliencia*. https://www.lamoncloa.gob.es/temas/fondos-recuperacion/Documents/30042021-Plan_Recuperacion_%20Transformacion_%20Resiliencia.pdf

Gobierno de España. (2021b). *Plan de Recuperación, Transformación y Resiliencia. Componente 29. Mejora de la eficacia del gasto público*. https://www.lamoncloa.gob.es/temas/fondos-recuperacion/Documents/16062021-Componente29.pdf

Gudiño, F. M. (2015). Evaluación y aprendizaje en la cooperación para el desarrollo: un análisis del caso español [tesis doctoral, Universidad Complutense de Madrid]. https://eprints.ucm.es/id/eprint/34477/1/T36737.pdf

Herrero, C. (2021). Gestión de recursos públicos y evaluación: un binomio inseparable. *Boletín Económico de ICE*, 3135, 23-35.

Instituto para la Evaluación de Políticas Públicas. (2021). *Estudio de Diagnóstico de la Evaluación de Políticas Públicas en la Administración General del Estado*. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Informes-de-Evaluacion/DIAGNOSTICO-EVALUACION-AGE/ESTUDIO_DIAGNOSTICO_EVA_AGE.pdf.pdf

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

Lázaro, B. (2015). *Comparative study on the institutionalisation of evaluation in Europe and Latin America*. Eurosocial Studies No. 15.

Meyer, W., Stockmann, R., & Taube, L. (2020). The Institutionalization of Evaluation. Theoretical Background, Analytical Concept and Methods. In R. Stockmann, W. Meyer & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 3-33). Palgrave Macmillan.

Munárriz, Á. (2020, 20 de septiembre). ¿Quién controla que las políticas sirven de algo? La falta de evaluación dispara el riesgo de una reconstrucción a ciegas. *Infolibre*. https://www.infolibre.es/noticias/politica/2020/09/18/quien_controla_que_politica_sirven_algo_falta_evaluacion_dispara_riesgo_una_reconstruccion_ciegas_111143_1012.html

OCDE. (2021). *Evaluation frameworks and practices: a comparative analysis of five OECD countries*.

Pattyn, V. (2012). *Why organisations (do not) evaluate. A search for necessary and sufficient conditions* (COMPASSS Working Paper No. 2012-70).

Pollitt, C., & Bouckaert, G. (2017). *Public Management Reform. A Comparative Analysis into the Age of Austerity*. Oxford University Press.

Rist, R. C., Boily, M. H., & Martin, F. (2011). *Influencing Change: Building Evaluation Capacity to Strengthen Governance*. World Bank. https://openknowledge.worldbank.org/handle/10986/2285

Rosenstein, B. (2015). *Status of national evaluation policies global mapping report (2.ª edición)*. Parliamentarians Forum on Development Evaluation in South Asia and EvalPartners. https://globalparliamentarianforum.files.wordpress.com/2016/02/the-status-of-evaluation-policies.pdf

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2003). *Evaluation: A Systematic Approach. Seventh Edition*. Sage Publications, Inc.

Ruiz, A. (2012). Panorámica actual de la evaluación de las políticas públicas. *Presupuesto y Gasto Público*, 68, 13-23.

Stockmann, R., Meyer, W., & Taube, L. (Eds). (2020a). *The Institutionalisation of Evaluation in Europe*. Palgrave Macmillan.

Stockmann, R., Meyer, W., & Taube, L. (2020b). The Institutionalisation of Evaluation in Europe: A Synthesis. In R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 483-521). Palgrave Macmillan.

Stufflebeam, D. L., & Shinkfield, A. J. (1995). *Evaluación sistemática: Guía teórica y práctica*. España: Centro de Publicaciones del Ministerio de Educación y Ciencia. Ediciones Paidós Ibérica.

Unión Europea. (2007). *Developing Evaluation Capacity*. https://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/report_integrated_2007.pdf

Vedung, E. (2010). *Four waves of evaluation diffusion. Evaluation, 16*(3), 263-277.

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Los retos de la explotación masiva de datos administrativos para la investigación y las políticas públicas: una visión desde dentro de la Administración

The challenges of the massive use of administrative data for research and public policy: a view from inside the public administration

Israel Arroyo
José Fernández Albertos
Ministerio de Inclusión, Seguridad Social y Migraciones

Resumen

Con el fin de contribuir al debate en torno a cómo corregir el déficit de explotación de datos administrativos para fines de investigación y evaluación de políticas públicas en el caso español, se analiza la cuestión desde la perspectiva de las administraciones que generan y gestionan esos datos, y se describen los principales obstáculos que están detrás de ese déficit. En la segunda parte del artículo se exponen una serie de cambios y reformas para revertirlos o sortearlos.

Palabras clave: *datos administrativos, big data, registros administrativos, evaluación de políticas, análisis de políticas públicas*

Códigos JEL: *C80, D63.*

Abstract

To contribute to the debate on how to correct for insufficient use of administrative data for research or policy evaluation in the Spanish case, we analyze the issue from the perspective of the administrations that produce and manage administrative records and describe the main obstacles that explain the deficit in its use. In the second part of the article we present a series of changes and reforms that could contribute circumventing these obstacles.

Keywords: *administrative data, big data, administrative records, policy evaluation, public policy analysis*

DOI: https://doi.org/10.32796/cice.2021.102.7309
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

La revolución que ha supuesto la posibilidad de explotar de forma masiva y entrelazada grandes bases de datos para la investigación y el diseño de políticas públicas en las últimas décadas ha dotado de un nuevo valor al papel de la administración como creador y gestor de registros administrativos. De hecho, no es exagerado afirmar que el «descubrimiento» de las posibilidades que ofrecían estos registros ha sido una de las principales fuerzas que han impulsado esta revolución.

La explotación de datos administrativos para la investigación académica y la evaluación de políticas públicas ofrece ventajas que van más allá de las asociadas al poder estadístico que confieren su alto número de observaciones. Así, a diferencia de otras formas de recogida de datos, el mantenimiento de registros administrativos permite el análisis de la información procedente de colectivos de difícil acceso (por ejemplo, es bien conocida la dificultad de los estudios basados en encuestas de ingresos o gastos para capturar información en los dos extremos de la distribución de la renta), y la combinación de fuentes de información de naturaleza diversa que hasta hace poco era difícil de integrar dentro de un mismo análisis. Igualmente, la regularidad con la que por lo general se mantienen los registros administrativos permite además una explotación longitudinal de la información muy difícil de replicar con otro tipo de fuentes.

La puesta a disposición de estas bases de datos para finalidades académicas y de evaluación y diseño de políticas públicas no es sin embargo un proceso sencillo (Perrin *et al*., 2015). En última instancia, el problema consiste en que estos datos administrativos se han recogido para mantener un recuento de operaciones de la propia administración, y no tanto (o no principalmente) para impulsar el análisis y la evaluación de las políticas que la administración despliega. Eso explica que la mayor parte de países de nuestro entorno hayan tenido que crear una nueva arquitectura legal e institucional para permitir el uso reglado y seguro de estos datos. En España hemos ido con algo de retraso en el desarrollo de esa arquitectura, lo que hace que todavía hoy la existencia de registros administrativos, generados y alimentados durante décadas para finalidades de gestión, coexiste con una explotación limitada y parcial de los datos contenidos en ellos. Es importante que nuestro país no se quede atrás en este cambio, y tenemos la ventaja de que podemos aprender de las estrategias de aquellos países que lo han hecho antes[1]. Pero además de importar las mejores prácticas, es preciso conocer bien los obstáculos con los que esta revolución se enfrenta en nuestro contexto.

Afortunadamente, existe un creciente interés público en torno a esta cuestión, que ha dado pie a un vibrante conjunto de iniciativas y contribuciones que están poniéndolo en el centro de la agenda política (Almunia & Rey-Biel, 2021; De la Fuente *et al*., 2021; AIReF, 2020). Con el fin de contribuir a este debate en torno a cómo corregir el déficit de explotación de datos en el caso español, en este artículo catalo-

[1] En AIReF (2020) o Brugarolas *et al*. (2020) se pueden encontrar algunos ejemplos.

gamos lo que a nuestro juicio son los principales obstáculos que están detrás de ese déficit, y a continuación exponemos una serie de cambios y reformas para revertirlos o sortearlos. En este abanico de medidas incluimos actuaciones ya realizadas en el ámbito de la Seguridad Social, una de las principales instituciones estatales generadoras de datos en nuestro país, pero también intervenciones de naturaleza más amplia de medio plazo que a nuestro juicio serán esenciales para consolidar un nuevo marco institucional de explotación de datos administrativos para la investigación.

2. La naturaleza multidimensional del problema: hacia una taxonomía de los obstáculos al uso de datos administrativos para la investigación

A la hora de entender cuáles son los principales pasos a dar para incorporarnos a la primera línea en la utilización de datos procedentes de registros administrativos con fines de investigación, es preciso hacer un ejercicio de identificación de los obstáculos que todavía hoy persisten. Estos obstáculos son de naturaleza diversa, y van desde problemas de diseño institucional a la existencia de diferentes tipos de inercias en el seno de las Administraciones públicas. La principal conclusión de este repaso es que no existe una explicación sencilla y unívoca a la pregunta de por qué los datos administrativos españoles están hoy infrautilizados, y el corolario es que tampoco existe una «bala de plata» que solucione este problema. Necesitaremos diferentes tipos de reformas e intervenciones.

2.1. El marco legal garantista y el problema de la confidencialidad y privacidad

En España, la legislación que regula la producción y uso de registros administrativos es por lo general muy garantista, y como principio no contempla expresamente el uso de los mismos para fines distintos a los que determinan su creación y en particular para fines de investigación o de evaluación. A falta de modificaciones normativas, la utilización ha de ampararse en los casos, muy tasados, que la normativa general de protección de datos contempla para fines de investigación.

La legislación reconoce el derecho de acceso a la información administrativa con fines estadísticos, pero no para investigación aplicada, aunque la legislación comunitaria (Reglamento (UE) 2016/679) sí contempla la cesión de la información en poder de la administración para la investigación científica, y se promueve la apertura de documentos y datos de las Administraciones públicas (Directiva 2019/1024 relativa a los datos abiertos y la reutilización de la información del sector público).

En términos sustantivos, el principal conflicto al que se enfrentan las administraciones a la hora de decidir si permitir o no el acceso de la comunidad investigadora a los datos administrativos que custodian es el de las implicaciones que esa cesión pueda tener en relación a la salvaguarda de la confidencialidad y la privacidad de los

individuos. Los datos administrativos no los recopilan las Administraciones públicas con fines investigadores –lo que crea problemas y por defecto, los titulares no han dado consentimiento expreso para que se usen con otros fines–. Esto, en ausencia de un marco legal que explícitamente establezca unas directrices para facilitar el uso de estos datos para la investigación o el análisis, empuja de manera natural a las administraciones a mantener una actitud en exceso recelosa a la hora de facilitar a terceros la explotación de estos datos.

2.2. La ausencia de expertise en protección de datos en el seno de las Administraciones públicas

Las restricciones que impone esta legislación garantista se magnifican por un problema más específico: la ausencia, por lo general, de auténticos expertos en materia de protección de datos en el seno de las Administraciones públicas. Ello conduce a una tendencia estructural hacia la realización de una interpretación siempre muy defensiva de la legislación de protección de datos por parte de las instituciones que mantienen las bases de datos administrativas.

Esta falta de *expertise* en protección de datos, añadido a una razonable y comprensible aversión al riesgo de los funcionarios responsables del mantenimiento de estas bases de datos tienden a desembocar en una socialmente indeseable tendencia a la opacidad. Igual que como dice el dicho, «nadie fue despedido nunca por comprar IBM», nadie se metió nunca en un lío por negarse a la utilización de unos datos para la investigación. Dicho de otra forma, los incentivos que tienen los responsables de estos ficheros son completamente opuestos a su utilización para investigación y en particular a su cesión para ese fin. En los casos en que existe ambigüedad, una interpretación favorable a la cesión de datos para el análisis puede conllevar responsabilidades para el funcionario o directivo público al cargo de los mismos, mientras que cuesta imaginar bajo qué circunstancias este mismo responsable pueda beneficiarse de las eventuales y siempre inciertas consecuencias beneficiosas de la cesión (en términos de mayor conocimiento científico, *inputs* de mayor calidad de cara a la elaboración de políticas públicas futuras, etcétera).

Este déficit de *expertise* no consiste solo en ausencia de perfiles profesionales especializados en diferentes disciplinas en protección de datos en el seno de las administraciones, sino que se manifiesta en las estructuras que gestionan estos registros. Al fin y al cabo, los «productores» de registros administrativos son fundamentalmente administraciones encargadas de lo que definimos genéricamente como «gestión». Son los encargados de recaudar los impuestos, gestionar el flujo de cotizaciones y prestaciones de la Seguridad Social, el funcionamiento del sistema sanitario, reducir la accidentalidad vial o asegurar el mantenimiento regular del padrón de población. El ejercicio de estas funciones en las sociedades modernas supone una gran especialización de las administraciones encargadas de llevarlas a cabo, que por lo general las realizan con una gran eficiencia.

Esta eficiencia y especialización «productiva», sin embargo, puede acabar reñida con la realización de otros fines necesarios para una buena gestión de las bases de datos, especialmente si esas bases han de generarse y mantenerse con los ojos puestos en la evaluación de agentes externos. La difusión de los registros es una tarea a menudo compleja, que conlleva la depuración de los registros, la interpretación y selección de los datos con la finalidad buscada de antemano, la creación de metadatos y la elaboración de documentos metodológicos, casos de uso… Tareas todas muy laboriosas e intensivas en mano de obra especializada ya no tanto en la naturaleza sustantiva de la política o programa en cuestión, sino en tratamiento y gestión de datos. La complejidad de estas se multiplica cuando además la evaluación –y las exigencias que impone a la hora de mantener los registros– no es tenida en cuenta en el diseño original de estos registros, lo que suele ser un problema especialmente en los registros de mayor longevidad.

Una de estas tareas que estas administraciones han de realizar es la de la aplicación de técnicas informáticas de anonimización que constituyen hoy un campo específico dentro de la ciencia de datos. Los avances en inteligencia artificial, que permiten incluso reconstruir datos de registro a partir de datos agregados, hacen que algunos métodos tradicionales para combinar la difusión de información con el mantenimiento del requisito de la anonimización se estén quedando obsoletos. Así, por ejemplo, puede resultar posible reconstruir una matriz original con identificadores por individuo a partir de distribuciones marginales procedentes de esa matriz en la que los identificadores no están presentes. Y las soluciones para mantener la anonimización ante estos problemas no son inocuas para el análisis (véase por ejemplo AEPD, 2019). Todo ello hace aún más problemático la infradotación de científicos de datos especializados en la gestión de estas bases administrativas.

En definitiva, todas estas tareas difícilmente pueden ser llevadas a cabo con eficacia en ausencia de recursos humanos y materiales específicos. Y en ausencia de cambios institucionales, no es realista pensar que las distintas administraciones productoras de registros vayan a dedicar parte de sus medios a una tarea que muchas veces no perciben como central dentro de sus funciones.

2.3. Fragmentación competencial

La ausencia de concentración de responsabilidades en una sola institución contribuye a veces a magnificar al problema anterior. No existe, a día de hoy, una asignación específica a ningún organismo no ya de las tareas anteriores sino ni siquiera de las funciones de evaluación que en última instancia podrían ejercer una labor tractora sobre la necesidad de impulsar las tareas mencionadas en el punto previo. Desaparecida la Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (AEVAL), cuyas competencias nunca fueron suficientes para llevar a cabo el cambio cultural e institucional necesario, existen a día de hoy una maraña de entidades con competencias directa o indirectamente relacionadas con la

evaluación o el diseño de políticas –cabe destacar entre ellas la Autoridad Independiente de Responsabilidad Fiscal (AIReF), que desde 2021 dispone de nueva División de Evaluación del Gasto Público y que gracias a unas amplias competencias formales de acceso a datos en poder de las Administraciones públicas ha sido capaz de generar el embrión de un *hub* de datos administrativos dedicados a la evaluación– sin que ninguna de ellas disponga de la capacidad suficiente para realizar esa tarea de impulso y coordinación.

Una vertiente de este problema es la existencia de una visión patrimonialista de los registros administrativos por parte de las distintas administraciones responsables. Los titulares de los ficheros adoptan en muchas ocasiones una actitud privativa sobre los datos, oponiéndose explícita o tácitamente a su utilización por otros. Este problema no es fácil, en la práctica, de distinguir del anterior pues el titular del fichero a menudo usa argumentos de carácter legal (de nuevo, valiéndose de la interpretación defensiva de un marco legal interpretable) para avalar su oposición a la cesión de los datos. Y la oposición a la cesión de una institución provoca reacciones de cadena en la misma dirección en las demás.

Evidentemente, la existencia de una pluralidad de administraciones y entidades gestoras de datos administrativos no es siempre un obstáculo: es muy posible que esta pluralidad también facilite el surgimiento y generalización de buenas prácticas. El reto es lograr establecer el marco que posibilite y potencie esta competencia y colaboración virtuosa mientras se minimizan los obstáculos que la fragmentación genera.

2.4. Debilidad de cultura de la evaluación

La debilidad de una cultura de la evaluación, entendida de modo genérico como creencia en que las políticas públicas han de estar informadas por la evidencia y que sus efectos han de ser analizados de acuerdo a los objetivos marcados[2], reduce los incentivos de los gestores y las administraciones a liberar y compartir datos cuando perciben que existe incertidumbre sobre el marco legal que ampara la distribución y explotación de datos.

Una versión especialmente perniciosa de este problema se produce cuando el titular del fichero se resiste a distribuir información no porque no incorpore en su cálculo coste-beneficio la contribución al conocimiento y a la mejora de la política pública que la explotación de datos puede generar, sino porque es intrínsecamente reacio a que se evalúe externamente la gestión de un programa o de una política.

Esta práctica es significativamente más problemática que la anterior, pues si los administradores eligen selectivamente qué ficheros liberar y cuáles no en base a los resultados esperados de la evaluación, ello acabará introduciendo un sesgo en los resultados de las evaluaciones que se realicen a partir de los datos cedidos.

[2] Un buen estudio empírico que aspira a medir la presencia de esta «cultura de la evaluación» a través de una encuesta a decisores públicos se puede encontrar en AIReF (2021).

Sin embargo, es importante señalar que el rechazo a liberar datos por las consecuencias de la evaluación puede tener otras muchas causas. Una posibilidad es que los gestores hayan desarrollado un interés intrínseco en el mantenimiento de una política o medida concreta, cuyo futuro pueda verse comprometido por una evaluación de los efectos de esa política. En este caso, el sesgo que esa actitud introduciría en las evaluaciones sería evidente. Pero también puede ser que un motivo por el cual las administraciones se opongan a liberar datos para la evaluación, más pedestre pero quizá más común, sea que el acceso por parte de terceros a esos registros puede revelar imperfecciones, errores, o sencillamente rutinas en la gestión de esos ficheros (a veces causado por la falta de recursos y de profesionales especializados en la gestión de estos ficheros), que los administradores prefieran mantener «fuera del radar».

Esto en el fondo revela una ventaja ulterior de la extensión de la cultura de la evaluación: la expectativa de que los datos almacenados en registros administrativos serán en algún momento analizados por evaluadores externos hará que los responsables de esos ficheros inviertan en mantenerlos en buen estado. Pero también muestra lo crucial que resulta la regulación y el ejercicio de las políticas de protección de datos: un acceso a los datos poco garantista debilitará los incentivos de los gestores a mantener datos con información rica, en la medida que la difusión indeseada de esa información pueda tener consecuencias comprometedoras para ellos en el futuro.

3. Medidas

3.1. Un marco legal proclive a la cesión de datos que dé seguridad jurídica a las administraciones

Como se ha señalado en la sección anterior, una buena parte de la reticencia de las administraciones a la cesión y uso de sus datos para usos de investigación o evaluación emana de la percepción que estos administradores tienen de la ausencia de amparo legal para avalar las decisiones de cesión de datos a terceros. Es, por consecuente, importante acomodar el marco legal que regula la gestión de los datos administrativos para que se incorpore de manera explícita este nuevo uso, con el fin de que las administraciones tengan un amparo claro y transparente para poder ceder datos, y que conozcan con claridad bajo qué condiciones de acceso pueden ceder estos ficheros.

Desde el Ministerio de Inclusión, Seguridad Social y Migraciones hemos promovido cambios legales en este sentido para permitir que los datos de la Seguridad Social, con las debidas y necesarias garantías, puedan ser utilizados por la comunidad investigadora. Así, el nuevo artículo 77 de la Ley General de Seguridad Social (modificado en el Real Decreto Ley 2/2021), que define explícitamente los fines para los cuales podrán ser usados los datos obtenidos por la Administración de la Seguridad Social en el ejercicio de sus funciones, incluye, por primera vez, en su apartado m):

> «fines de investigación científica en el ámbito de la protección social, en el marco establecido por el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos), incluidas las posibles comunicaciones instrumentales que, a efectos de la realización de la investigación, resulte preciso efectuar a sujetos distintos de aquellos que lleven a cabo directamente dicha investigación. Se entenderán comprendidas en esta finalidad las actividades de evaluación de las políticas públicas en materia de protección social.
>
> Los tratamientos que se efectúen en relación con esta finalidad se limitarán a los datos estrictamente imprescindibles para la realización de la actividad de que se trate, utilizándose los procedimientos adecuados que no permitan la identificación de los interesados. Ello no impedirá la comunicación de datos sin anonimizar a efectos meramente instrumentales cuando ello resulte imprescindible para realizar la actividad, se limite a los datos estrictamente necesarios, se garantice que el encargado del tratamiento no podrá utilizarlos con otra finalidad y el tratamiento ulterior garantice la no identificación de los interesados.»

Por supuesto, esto es solo un avance parcial, que afecta a solo uno de los proveedores de bases administrativas. Pero creemos que es un paso importante. Si se logra consolidar el hacer explícito en la regulación de gestión de datos la necesidad de facilitar su uso para la investigación y la evaluación, no solo se construirá un marco legal que dé certidumbre a los gestores y se haga más fácil y más seguro la explotación de estos datos para estos fines, sino que también se facilitará la movilización de recursos para eliminar los demás obstáculos identificados en la sección anterior.

En un plano más general, y de cara a la universalización de estas prácticas, las recomendaciones de cambios normativos señalados en AIReF (2020) para la creación de una institucionalización de estas prácticas son, creemos, el necesario punto de partida.

3.2. Creación de nuevas infraestructuras estables para la cesión y explotación de datos

En primer lugar, un cambio necesario es la incorporación de más perfiles especialistas en gestión y difusión de datos en el seno de las instituciones que gestionen datos administrativos. Las nuevas exigencias de la cesión para uso investigador y de evaluación de políticas exige de los administradores de estas bases de datos unas *skills* y una especialización diferente a las tradicionales, más centradas en los usos internos y de gestión que tradicionalmente se han dado a estos datos.

Pero más allá de este cambio en los perfiles, creemos que es necesario también cambiar las infraestructuras que permitan compaginar difusión y explotación con fines de investigación con la necesaria protección de la confidencialidad de los datos. En la Seguridad Social, un paso que hemos dado ha sido la puesta en marcha de «salas seguras» en el seno de instalaciones propias (hasta la fecha tres, situadas en diferentes partes del territorio) que permitan a investigadores externos, con las debidas garantías, el acceso a datos internos que por su naturaleza no puedan ser difundidos.

Las salas seguras son una forma de permitir el acceso a datos identificables (y por tanto muy sensibles) a cambio de un fuerte mecanismo de protección de esos datos. Son un paso importante, pero para maximizar el impacto de la explotación de estos datos es necesario poder cruzar ficheros procedentes de diferentes administraciones. Como el proceso de cruce introduce todavía mayores vulnerabilidades en términos de privacidad y confidencialidad, es importante acompañar ese proceso con la creación de infraestructuras que permitan combinar accesibilidad para la investigación con procesos más complejos y garantistas de anonimización de los datos.

La buena noticia es que las tecnologías para combinar protección de la anonimización con el cruce de datos procedentes de diferentes instituciones existen. En AIReF (2020) se hace un repaso de ellas. En Francia, por ejemplo, el Centro de Acceso Seguro a Datos (CASD) es un organismo creado a partir del Instituto Nacional de Estadística (INSEE). Este organismo actúa como «tercero seguro» frente a las distintas administraciones, que seguirían sin tener acceso a los identificadores de los registros de las demás, y que se encargaría de unir ficheros a través de identificadores y «devolverlos» de forma anonimizada a sus usuarios finales. Esta institución tampoco realiza investigación propia, para evitar un potencial conflicto de intereses. Igualmente, la decisión de facilitar el acceso a un determinado conjunto de datos se hace en función de peticiones concretas, que son evaluadas por un comité diferenciado que no se encarga de las tareas de cesión y depuración de los ficheros.

Para avanzar en esa dirección, el camino más razonable debería ser el de profundizar en el grupo de trabajo estable entre los principales productores de datos en España: el Instituto Nacional de Estadística, la Agencia Tributaria, el Servicio Público de Empleo Estatal, y la Seguridad Social, y que podría ser el germen de esa eventual nueva estructura. Este nuevo «consorcio» estable debería también contar con la participación de la Agencia Española de Protección de Datos (AEDP), que actuaría como garante último de que los procesos de explotación sean totalmente compatibles con el mantenimiento de la confidencialidad y anonimidad de los datos.

Esta institucionalización de la colaboración inter-institucional también tendría otras consecuencias saludables, al suponer un impulso a la estandarización de criterios para identificar individuos u otras unidades de análisis (empresas, entidades, hogares…) en diferentes bases de datos, compartir el *expertise* generado dentro de cada una de estas instituciones de manera transversal entre ellas, y el aprovechar las economías de escala que se generarían por cada fusión de datos, al ser estas fusiones explotables por investigadores diferentes a los que hicieron la petición original.

4. Conclusiones

En última instancia, el éxito del proceso de apertura de datos administrativos para la investigación y el análisis dependerá de si estas u otras intervenciones acaban generando una fuerte cultura de la evaluación y del análisis en la comunidad académica e investigadora, pero también con capacidad de permear el día a día de las Administraciones públicas.

España parte con algo de desventaja en la explotación de las bases de datos administrativas para mejorar la elaboración de políticas públicas, pero creemos que la toma de conciencia de los obstáculos, las iniciativas que se están desplegando, el aprendizaje a partir de la experiencia reciente de los países en nuestro entorno, y la cada vez mayor concienciación tanto en la sociedad como en las instituciones acerca del gran potencial que estos cambios pueden generar para el bienestar común son motivos para ser razonablemente optimistas.

Referencias bibliográficas

AEPD, Agencia Española de Protección de Datos. (2019). *La k-anonimidad como medida de la privacidad*. Nota técnica 0/2019. https://www.aepd.es/sites/default/files/2019-09/nota-tecnica-kanonimidad.pdf

AIReF, Autoridad Independiente de Responsabilidad Fiscal. (2020, octubre). *Opinión para una estrategia de acceso a datos administrativos*. Opinión 1/2020. https://www.airef.es/wp-content/uploads/2020/10/OPINION_DATOS/201005.-Opinion-Acceso-Datos.pdf

AIReF, Autoridad Independiente de Responsabilidad Fiscal. (2021). *La institucionalización de la evaluación de políticas públicas en Castilla y León: situación actual y propuestas*. Estudio de Evaluación de Políticas. https://www.airef.es/wp-content/uploads/2021/10/ESTUDIOS/PDF-WEB-accesible-CASTILLA-Y-LEON-1.pdf

Almunia, M., & Rey-Biel, P. (2020). *Por un cambio de cultura en la gestión de los datos en España: Una propuesta de reforma*. EsadeEcPol Insight, 17/2020. https://www.airef.es/wp-content/uploads/2020/10/OPINION_DATOS/201005.-Opinion-Acceso-Datos.pdf

Brugarolas, P., De la Rica, S., & Gorjón, L. (2020). *El uso de datos administrativos para la investigación. El caso de Dinamarca: un ejemplo de buenas prácticas*. Fundación ISEAK. https://iseak.eu/wp-content/uploads/2020/11/El-sistema-danés-demicrodatos_DFG.pdf

De la Fuente, A., de Rus, G., Fernández M., García M. A., Jansen M., Jiménez S., Novales A., Onrubia J., Pérez Renovales J., Sastre E., & Sicilia J. *La evaluación de políticas públicas en España: antecedentes, situación actual y propuestas para una reforma*. FEDEA Policy Papers 2021/09. https://documentos.fedea.net/pubs/fpp/2021/10/FPP2021-09.pdf?utm_source=wordpress&utm_medium=portada&utm_campaign=estudio

Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público (versión refundida). *Diario Oficial de la Unión Europea*.
http://data.europa.eu/eli/dir/2019/1024/oj

Real Decreto-ley 2/2021, de 26 de enero, de refuerzo y consolidación de medidas sociales en defensa del empleo. *Boletín Oficial del Estado*, n.º 23, de 27 de enero de 2021.
https://www.boe.es/eli/es/rdl/2021/01/26/2/con

Reglamento (UE) 2016/679 del Parlamento europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). *Diario Oficial de la Unión Europea*.
http://data.europa.eu/eli/reg/2016/679/oj

Perrin, S., Barrigar J., & Gellman R. (2015). *Government Information Sharing Is Data Going Out of the Silos, Into the Mines?*
https://www.oipc.ab.ca/media/389571/Report_Government_Information_Sharing_Jan2015.pdf

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Estadística oficial y evaluación de políticas públicas: oportunidades y retos

Official statistics and public policy evaluation: opportunities and challenges

Juan Manuel Rodríguez Póo
Antonio Salcedo Galiano
Instituto Nacional de Estadística

Resumen

La estadística oficial constituye un bien público al servicio de las sociedades democráticas. La pandemia de la COVID-19 ha demostrado la extraordinaria resiliencia de la estadística, aunque también ha puesto de manifiesto la necesidad de adoptar medidas adicionales orientadas a establecer un mejor uso de las nuevas fuentes de datos, que ofrecen una oportunidad única a la estadística para satisfacer de forma eficiente las crecientes demandas de los usuarios, incrementando la frecuencia y granularidad de la información así como la disponibilidad de nuevos conjuntos de datos, obteniendo una estadística de alta definición en tiempo real que redundaría en un mejor análisis y evaluación de políticas públicas.

Palabras clave: *big data, código de buenas prácticas, estadística oficial, evaluación de políticas públicas.*

Clasificación JEL: *C00, H11.*

Abstract

Official statistics constitute a public good at the service of the democratic societies. The COVID-19 pandemic has shown the extraordinary resilience of statistics, although it has also highlighted the need to take additional measures aimed at making a better use of new data sources, which offer a unique opportunity for statistics to efficiently meet the growing demands of users, increasing the frequency and granularity of the information as well as the availability of new data sets, obtaining high definition statistics in real time that would result in a better analysis and evaluation of public policies.

Keywords: *big data, code of good practices, official statistics, public policy evaluation.*

DOI: https://doi.org/10.32796/cice.2021.102.7310
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

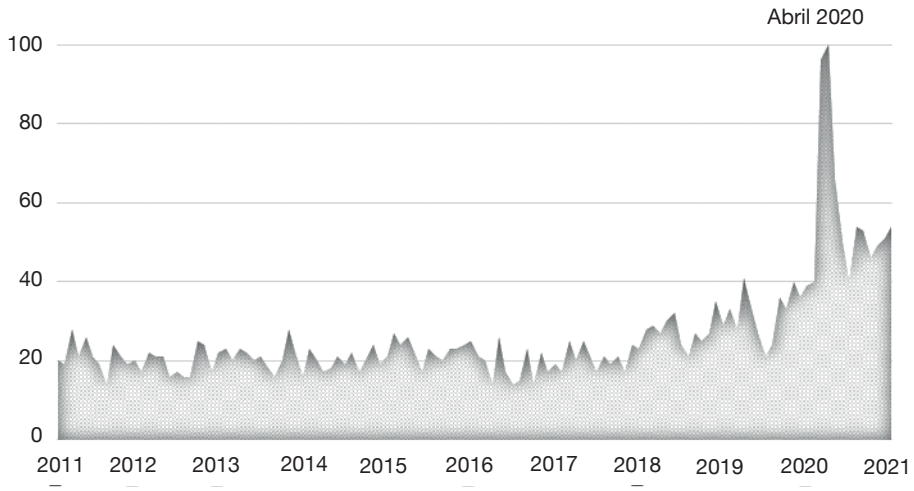
La información estadística constituye un elemento de gran importancia en el proceso de evaluación de políticas públicas, aportando la evidencia empírica necesaria para ofrecer una visión precisa de la situación actual, así como para abordar la medición de las relaciones existentes entre las distintas políticas aplicadas y los efectos finalmente alcanzados (Feinstein, 2007). La cada vez mayor utilización de estadísticas en la evaluación de políticas proporciona un singular sustento al proceso de toma de decisiones, reforzando a su vez otros procesos relacionados tales como la transparencia o la rendición de cuentas (Fano, 2017). Este uso cada vez más intensivo de datos estadísticos para evaluación ha ido aparejado con el desarrollo de nuevas técnicas cuantitativas, sistematizadas según los distintos tipos de intervenciones públicas (Moral *et al*., 2016), junto con la aparición de novedosas formas de acceso a datos y presentación de resultados (UNECE, 2009).

La situación provocada por la pandemia de la COVID-19 ha impulsado aún más la fuerte demanda de estadísticas por parte de los evaluadores de políticas públicas y, de forma general, por la sociedad, en diferentes ámbitos: sociodemográfico, económico, sanitario o de movilidad. Considerando el periodo de tiempo de diez años que abarca desde el 1 de enero de 2011 al 1 de enero de 2021, en la Figura 1 puede apreciarse que los valores máximos de la tendencia de búsqueda de «estadísticas» en España se obtuvieron entre los meses de marzo y abril del año 2020, fechas que precisamente coinciden con el punto álgido de la pandemia. Desde entonces este índice de búsquedas se mantiene en un valor más elevado que la media de años previos.

Sin duda la COVID-19 ha supuesto un extraordinario desafío para la estadística oficial, cuya producción se basaba en buena medida en modelos clásicos de recogida de datos a partir de encuestas (Statistics Canada, 2003), evidenciando la aparición de nuevos y desafiantes retos tecnológicos para la estadística, desde la recogida de datos hasta la propia difusión de la información. Los sistemas estadísticos nacional y europeo han afrontado una crisis sin precedentes y han buscado dar respuesta a dicho desafío, poniendo en marcha nuevos productos estadísticos y publicando información orientada a satisfacer las nuevas demandas de datos, en ocasiones casi en tiempo real.

En este contexto el uso intensivo de los registros administrativos y de forma más general, el *big data*, han irrumpido con extraordinaria fuerza en la estadística oficial, aportando también importantes beneficios al diseño de políticas públicas, lo que a su vez ha redundado en una mejora de la calidad y la eficiencia de los servicios públicos así como en la más eficiente utilización de recursos (Azzone, 2018). No obstante, el tránsito desde esos grandes volúmenes de datos hacia un conjunto estable de indicadores sintéticos es un camino arduo y complejo que presenta un buen número de dificultades (Arnaboldi & Azzone, 2020), la primera de las cuales es el propio acceso a esas nuevas fuentes de datos que, en múltiples ocasiones, se encuentran mantenidas por poseedores o tenedores privados de datos.

FIGURA 1
ÍNDICE DE BÚSQUEDA DE «ESTADÍSTICAS» EN ESPAÑA ENTRE LOS AÑOS 2011-2021



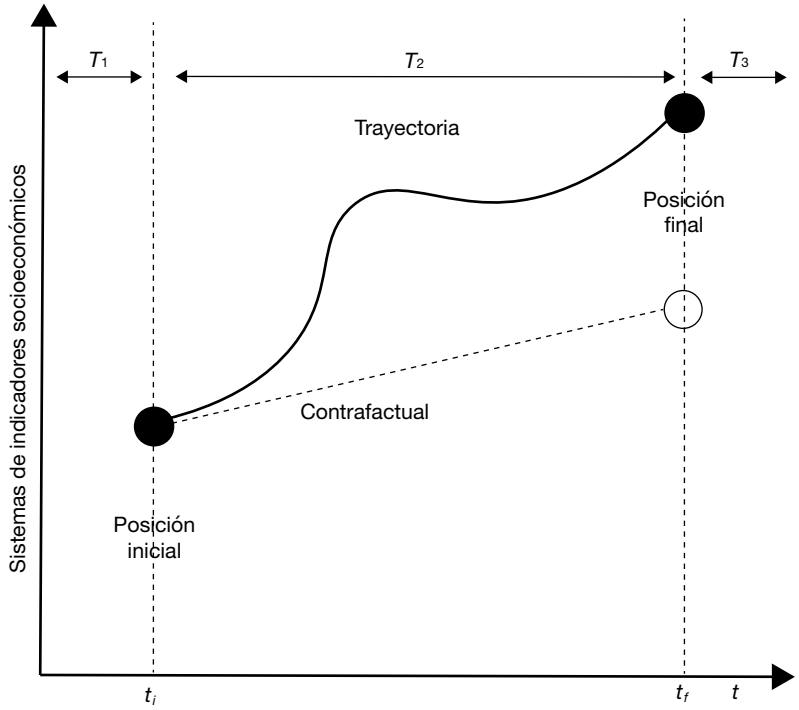
FUENTE: Google Trends y elaboración propia.

En una sociedad altamente datificada y globalizada la estadística oficial debe desempeñar un papel clave en la toma de decisiones basadas en datos (*data-driven decision-making*) o, de forma más amplia, en el uso de datos para políticas (*data for policy, D4P)* (Radermacher, 2019). Este rol requiere un estudio detallado de la situación de la estadística oficial en la actualidad bajo la premisa de que este bien público tenga asegurado que podrá disponer de una posición relevante en la definición de las políticas de datos en el corto, medio y largo plazo, facilitando así la continuidad de su desempeño con la máxima eficacia.

2. Estadísticas, indicadores y evaluación de políticas públicas en España

Un indicador estadístico constituye una medida resumida relacionada con un problema o fenómeno clave y que se deriva de una serie de hechos observados (Eurostat, 2014). Los conjuntos o sistemas de indicadores constituyen una herramienta clásica a la hora de acometer evaluaciones de políticas públicas. Estos sistemas nos aportan tanto una foto fija de la actuación pública en estudio como su evolución, calculados a partir de los distintos resultados obtenidos. Desde una óptica temporal una política pública puede representarse de forma simplificada de acuerdo con la Figura 2. Esta política pública vendría caracterizada por una posición inicial en un momento t_i, una trayectoria determinada por los resultados en cada instante temporal t del conjunto de indicadores socioeconómicos seleccionados y, por último, una posición final que sucede en el momento t_f (Pollitt & Bouckaert, 2010).

FIGURA 2
EVOLUCIÓN DE INDICADORES SOCIOECONÓMICOS DE UNA POLÍTICA PÚBLICA



FUENTE: Elaboración propia.

Estos elementos básicos a su vez nos permiten diferenciar tres periodos de tiempo (T_1, T_2 y T_3) en los que habitualmente se acometen las evaluaciones de las políticas públicas, ya sean *ex-ante*, intermedia, *ex-post* o continua. De este modo los resultados de los indicadores seleccionados nos permiten disponer de una medida del fenómeno estudiado y nos revelan posiciones relativas, cambios o tendencias, lo que además puede compararse con una situación contrafactual que podría haber sucedido en caso de que no se hubiera aplicado la actuación pública (Loi & Rodrigues, 2012). Esto permite, de forma restrospectiva, abordar análisis situacionales (¿qué pasó?) o de causa-efecto (¿por qué pasó? ¿cuál fue la causa más probable?) así como, de forma prospectiva, diferentes estudios predictivos (¿qué pasará?) o prescriptivos (¿qué pasaría?). Claramente cada proceso de evaluación queda condicionado por la cantidad de datos -y microdatos- disponibles, los periodos de referencia del conjunto de indicadores utilizado y, de forma más general, con el grado de calidad de la información que proporcionan.

En nuestro país la estadística oficial tiene una tradición muy amplia que se remonta al año 1856, con la creación de la Comisión de Estadística del Reino (Merediz, 2004), si bien con anterioridad se habían elaborado distintos trabajos censales de gran interés como el censo de pecheros de 1528, el censo de la Corona de Castilla de 1591 o el censo de Floridablanca de 1787. En sus orígenes la estadística estaba más estrechamente vinculada con los asuntos del Estado (Barreto-Villanueva, 2012) como, por ejemplo, el pago de impuestos, el reclutamiento de soldados o el reparto de tierras u otros bienes. Actualmente la producción de estadísticas se basa en una serie de principios (Eurostat, 2018) establecidos en el Código Europeo de Buenas Prácticas (CEBP), que proporcionan un sello de calidad único que facilita que la estadística mantenga un alto grado de credibilidad y confianza por parte de los usuarios. Estos conjuntos de principios de la estadística oficial se muestran en el Cuadro 1.

El grado de cumplimiento de todos esos principios tiene, sin duda, efectos en la realización de evaluaciones de políticas públicas. Así, desde una óptica temporal, la pertinencia, oportunidad, puntualidad, coherencia y comparabilidad de resultados constituyen una serie de propiedades requeridas a los indicadores estadísticos que tengan que ser utilizados en la evaluación. Por su parte, desde una óptica de las fuentes de datos, resulta necesario que estas sean independientes, confiables, basadas en una metodología sólida, con procedimientos estadísticos adecuados, que eviten una carga excesiva para las unidades informantes y sean rentables en el sentido de acometer un uso eficiente de los recursos. Desde el punto de vista de la política que se pretenda explicar, resulta especialmente significativo que las estadísticas se elaboren con una frecuencia y periodicidad suficiente, aportando además un adecuado nivel de desagregación.

En España los datos estadísticos oficiales constituyen un input necesario y fundamental para la evaluación de políticas públicas de organismos como la Autoridad Independiente de Responsabilidad Fiscal (AIReF, 2020a) o el Instituto para la

CUADRO 1

PRINCIPIOS DEL CEBP APLICADOS A LA ESTADÍSTICA OFICIAL DE ESPAÑA

Entorno institucional	Procesos estadísticos	Producción estadística
1. Independencia profesional. 1. (bis). Coordinación y cooperación. 2. Mandato de recogida de datos y acceso a los datos. 3. Adecuación de los recursos. 4. Compromiso de calidad. 5. Confidencialidad estadística y protección de datos. 6. Imparcialidad y objetividad.	7. Metodología sólida. 8. Procedimientos estadísticos adecuados. 9. Carga no excesiva para los encuestados. 10. Rentabilidad.	11. Pertinencia. 12. Precisión y fiabilidad. 13. Oportunidad y puntualidad. 14. Coherencia y comparabilidad. 15. Accesibilidad y claridad.

FUENTE: Instituto Nacional de Estadística (INE) y elaboración propia.

Evaluación de Políticas Públicas (Rico, 2020). La estadística oficial española se regula por medio de la Ley 12/1989, de 9 de mayo, de la Función Estadística Pública (LFEP), que normaliza la función estadística para fines estatales, la planificación y elaboración de las estadísticas desarrolladas por la Administración del Estado, y la organización de sus servicios estadísticos y sus relaciones en materia estadística con las comunidades autónomas, las corporaciones locales, la Unión Europea y los organismos internacionales. La LFEP establece el Plan Estadístico Nacional (PEN) como el principal instrumento ordenador de la actividad estadística de la Administración General del Estado. El PEN contiene las estadísticas que han de elaborarse a lo largo de un cuatrienio por los servicios estadísticos del Estado y las que han de llevarse a término, total o parcialmente, con participación de las comunidades autónomas y las corporaciones locales en virtud de acuerdos de cooperación o, en su caso, en ejecución de lo previsto en las leyes. Todas las estadísticas incluidas en el Plan tienen la consideración de estadísticas para fines estatales, se implementan mediante Programas anuales de desarrollo y son de cumplimentación obligatoria.

Si bien la ley estadística de 1989 ha permitido gobernar eficazmente la estadística oficial durante más de treinta años, actualmente existe un amplio consenso en que su revisión resultaría conveniente para poder adaptarla a las nuevas posibilidades de información que ofrece el actual proceso de transformación digital (Consejo Superior de Estadística, 2018).

3. Estadísticas y evaluación de las políticas públicas en el contexto europeo

En la Unión Europea (UE) las evaluaciones de políticas públicas de la Comisión siguen los principios de la iniciativa *Better Regulation* (Listorti *et al*., 2020). La agenda de *Better Regulation* está orientada hacia una evaluación sistemática del impacto económico, social y ambiental de la acción política para, así, tratar de asegurar que la legislación propuesta sea del más alto nivel. Un pilar para conseguir una mejor regulación es la evidencia empírica, que incluye la información cuantitativa y cualitativa proporcionada por la estadística oficial como elemento clave para describir el problema, comprender mejor su causalidad y, consecuentemente, la lógica de la intervención pública, y finalmente poder evaluar el impacto alcanzado. Por tanto, desde la óptica de la UE resulta fundamental que las estadísticas europeas dispongan de un fuerte grado de armonización y comparabilidad, tanto en la dimensión temporal (años, trimestres, meses) como espacial (estados miembros, regiones, ciudades), lo que a su vez permite y facilita la construcción de agregados europeos.

La primera «ley de estadística europea» (CE, 1997) se aprobó poco después de la inclusión del artículo 213-A en el Tratado de Ámsterdam, que habilitaba la elaboración de estadísticas comunitarias «cuando sean necesarias para la realización de las actividades de la Comunidad». Esta primera ley estadística fue remplazada, doce años después, por un nuevo reglamento sobre estadísticas europeas (CE, 2009) que sería revisado más tarde, en el año 2015. Todas estas normas han habilitado el desarrollo

de diversos reglamentos estadísticos sectoriales, que garantizan una producción estadística armonizada en múltiples áreas temáticas, entre otras, las estadísticas coyunturales y estructurales de empresas, las estadísticas de precios y mercado laboral, las estadísticas sociales, o las estadísticas de agricultura y medio ambiente. También han facilitado la adopción de los sistemas contables europeos –incluida la contabilidad nacional y las cuentas satélite– y de varios conjuntos de indicadores europeos (Euroindicadores, Objetivos de Desarrollo Sostenible, Pilar Europeo de los Derechos Sociales, Economía Circular y Procedimiento de Desequilibrio Macroeconómico) para monitorizar distintas políticas públicas conforme se reflejan en el Cuadro 2.

CUADRO 2
SISTEMAS DE INDICADORES EN EL ÁMBITO DE LAS ESTADÍSTICAS EUROPEAS

Conjunto de indicadores	Ámbitos
Euroindicadores	– Balanza de pagos. – Encuestas a empresas y consumidores. – Industria y servicios. – Comercio internacional. – Mercado laboral. – Indicadores monetarios y financieros. – Cuentas nacionales. – Precios.
Objetivos de desarrollo sostenible	– Fin de la pobreza. – Hambre cero. – Salud y bienestar. – Educación de calidad. – Igualdad de género. – Agua limpia y saneamiento. – Energía asequible y no contaminante. – Industria, innovación e infraestructura. – Reducción de las desigualdades. – Ciudades y comunidades sostenibles. – Producción y consumo responsables. – Acción por el clima. – Vida submarina. – Vida de ecosistemas terrestres. – Paz, justicia e instituciones sólidas. – Alianzas para lograr los objetivos.
Pilar europeo de derechos sociales	– Igualdad de oportunidades y acceso al mercado laboral. – Mercados laborales dinámicos y condiciones laborales justas. – Apoyo de las autoridades públicas / Protección e inclusión social.
Economía circular	– Producción y consumo. – Gestión de residuos. – Materias primas secundarias. – Competitividad e innovación.
Procedimiento de desequilibrio macroeconómico	– Desequilibrios externos y competitividad – Desequilibrios internos. – Indicadores de empleo

FUENTE: Eurostat y elaboración propia.

Las estadísticas europeas, como servicio público, también son objeto de evaluaciones periódicas que refuerzan el fuerte compromiso existente con la calidad de los datos de los servicios estadísticos. Así en el Sistema Estadístico Europeo (SEE) se han definido cuatro niveles de garantía de calidad: el primero, formado por el conjunto de dieciséis principios recogidos en el Código Europeo de Buenas Prácticas; el segundo consiste en un conjunto de ochenta y cuatro indicadores relativos a cada uno de los principios anteriores; el tercero se basa en el denominado Marco de garantía de calidad del SEE (Eurostat, 2012), que proporciona diversos métodos y herramientas para facilitar la implementación práctica del Código de prácticas; en el cuarto nivel la calidad de los procesos y productos estadísticos individuales se evalúa en función de una serie de reglas estandarizadas e informes estructurados (Eurostat, 2020), que proporcionan a los usuarios información relevante sobre los métodos de recopilación y validación de datos tanto a nivel nacional como europeo. Estos cuatro niveles de garantía de la calidad se complementan con evaluaciones externas que, entre otros, incluyen un análisis periódico del grado de cumplimiento del Código en cada estado miembro por parte de Eurostat mediante los denominados *Peer Reviews*, los ejercicios de verificación de la Comisión Europea, u otras actuaciones específicas realizadas por terceros como, por ejemplo, el Tribunal de Cuentas Europeo (ECA, 2021).

La adopción del Reglamento 223/2009 sobre estadísticas europeas supuso la aprobación de provisiones específicas y novedosas que han permitido mejorar la calidad de los servicios proporcionados por la estadística oficial. En este punto resulta conveniente destacar dos de ellas establecidas en los artículos 17 bis y 23 de este Reglamento, respectivamente:

a) Las autoridades nacionales estadísticas «tendrán derecho a acceder a todos los registros administrativos sin demora y gratuitamente, a hacer uso de ellos y a integrarlos en las estadísticas en la medida necesaria para el desarrollo, la elaboración y la difusión de las estadísticas europeas». Asimismo se consultará sobre «el diseño inicial, el desarrollo posterior y la supresión de registros administrativos creados y mantenidos por otros organismos». Además, el Instituto Nacional de Estadística (INE) deberá «participar en las actividades de estandarización relativas a registros administrativos que sean de utilidad para la generación de estadísticas europeas».
b) Eurostat y los institutos nacionales de estadística «en sus respectivas esferas de competencia, podrán conceder el acceso a datos confidenciales que solo permitan la identificación indirecta de unidades estadísticas a investigadores que lleven a cabo análisis estadísticos con fines científicos», dando de este modo respuesta a la fuerte demanda existente para el acceso a los microdatos de las operaciones estadísticas (Crato & Paruolo, 2019).

Sin duda ambas medidas han supuesto un paso importante en la dirección de mejorar la calidad de las estadísticas, facilitando un uso más intensivo de los recursos estadísticos por parte de investigadores y salvaguardando en todo caso la confidencialidad y el secreto estadístico. Sin embargo, la legislación estadística de los países de la Unión Europea en general no contiene provisiones directas sobre evaluación de políticas públicas si bien podemos encontrar algunas referencias indirectas que podrían ser entendidas como una finalidad propia de las estadísticas. A modo de ejemplo, en Países Bajos se atribuye al Central Bureau of Statistics la misión de realizar actividades estadísticas con fines de investigación y publicar las estadísticas compiladas sobre la base de dicha investigación; por su parte la ley estadística de Letonia incluye entre los objetivos de la estadística oficial el monitoreo o evaluación de leyes y reglamentos o documentos de planificación del desarrollo, e incluso la redacción de los escenarios de previsiones; en Rumanía, su ley establece que la estadística oficial facilita la formulación, implementación, modernización y evaluación de políticas nacionales, planes de desarrollo nacionales y locales, de conformidad con los principios fundamentales de las estadísticas oficiales; y en Reino Unido, el Statistics Board tiene atribuida la función de apoyar en el desarrollo de evaluación de políticas públicas. Precisamente este último es el enfoque que se sigue en la ley genérica de estadística, que señala que debería atribuirse al consejo estadístico la tarea de «promover el uso de estadísticas oficiales como evidencia para la formulación de políticas públicas, su evaluación y toma de decisiones» (UNECE, 2018).

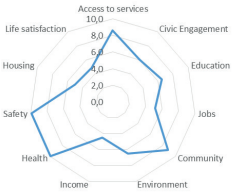
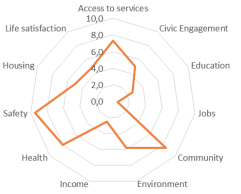
Por otra parte, las oficinas de estadísticas europeas son conscientes de las posibles soluciones que aportan los nuevos roles de administración y gestión de datos, de forma que se están produciendo adaptaciones en las funciones desempeñadas hasta ahora. Así, el Instituto Nacional de Estadística de Italia cuenta desde diciembre 2020 con el «Centro ISTAT para estadísticas inteligentes confiables» (Istat Centre for Trusted Smart Statistics) destinado a consolidar y desarrollar la experiencia interna en la gestión del *big data*, incluido el apoyo de la estadística oficial a cuestiones políticas de nivel gubernamental. En Lituania, tras la reciente reforma de su ley de estadística en octubre de 2020, la oficina de estadística cuenta con dos nuevas funciones: a) establecer el «Sistema de Información de Gobernanza de Datos Estatales»; b) gestionar la información contenida en el sistema de información del Estado. Por su parte, la Oficina Federal de Estadística de Suiza ejerce el liderazgo en la «Gestión nacional de datos» (*Nationale Datenbewirtschaftung, NaDB*) a través de una plataforma de interoperabilidad en la que se integran datos administrativos del gobierno central, de los cantones y de empresas privadas.

Situándonos en un marco geográfico más amplio, en Estados Unidos el acceso directo y seguro a los microdatos administrativos también se ha considerado una prioridad (Card *et al*., 2010). En cuanto a la Organización para la Cooperación y Desarrollo Económicos (OCDE) la estadística también constituye un elemento fundamental a la hora de elaborar mejores políticas para mejores vidas (*better policies for better lives*) así como para proceder a su evaluación (OCDE, 2020b). Esta organi-

zación internacional ha sido clave en el desarrollo de diversos productos estadísticos orientados para la evaluación de políticas públicas, entre otros, el Manual de Frascati para encuestas de I+D, las pruebas del programa para la evaluación internacional de alumnos (PISA) o el índice de calidad de vida de la iniciativa «How's life?» (OCDE, 2020a). Conforme se observa en la Figura 3, esta iniciativa proporciona estimaciones de bienestar nacional y regional considerando hasta once dimensiones diferentes: vivienda, ingresos, empleo, comunidad, educación, medio ambiente, compromiso cívico, salud, satisfacción, seguridad, balance vida-trabajo, lo que permite la realización de comparaciones transnacionales.

FIGURA 3

ESTIMACIÓN DEL BIENESTAR REGIONAL

Comunidad Autónoma	Regiones con similar nivel de bienestar en otros países
Cantabria	– Border, Midland and Western (Irlanda) – Occitanie (Francia) – Basilicata (Italia) – Auckland (Nueva Zelanda)
Madrid 	– Trento (Italia) – Provence-Alpes-Côte d'Azur (Francia) – Greater London (Reino Unido) – Berlin (Alemania)
Andalucía 	– Basilicata (Italia) – West Greece (Grecia) – Hauts-de-France (Francia) – Border, Midland and Western (Irlanda)

FUENTE: OCDE regional well-being.

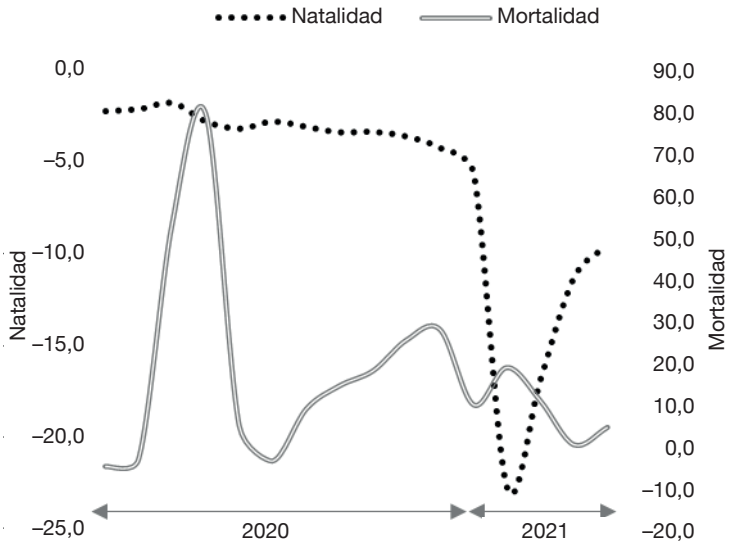
4. Hacia un nuevo paradigma de datos e información estadística para la evaluación de políticas públicas

En la actualidad vivimos en una sociedad altamente datificada que se caracteriza por la existencia de nuevos y variados ecosistemas de datos, que constituyen en sí una clara oportunidad para que la estadística oficial refuerce su estrecho vínculo con la evaluación de políticas públicas, aportando mucha más información puntual y granular. En efecto, el acceso a nuevas fuentes de datos permite avanzar en la puesta en marcha de nuevos productos estadísticos que, utilizando metodologías innovadoras, facilita dar respuestas a las necesidades de los investigadores y evaluadores de políticas públicas ofreciéndoles un desglose de datos mucho más detallado e incrementando la frecuencia de la disponibilidad de la información, es decir, nuevas estadísticas en alta definición y tiempo real.

Un claro ejemplo de esta nueva situación lo constituyen las estadísticas experimentales que han puesto en marcha los sistemas estadísticos nacional y europeo para, así, dar respuesta de manera más oportuna a las necesidades de los usuarios. Si bien estas estadísticas no necesariamente cumplen por el momento con todos los requisitos de calidad y grado de armonización exigibles a las estadísticas oficiales, su utilidad es manifiesta al cubrir claros vacíos de información estadística, facilitando el diseño de nuevos indicadores y proporcionando nueva información para el diseño y evaluación de políticas públicas. Existen varios prototipos de estadísticas experimentales, entre los cuales se pueden citar: la estimación del número de defunciones semanales durante el brote de la COVID-19, la estimación mensual de nacimientos; la coyuntura demográfica de empresas, el atlas de distribución de renta de los hogares, los estudios de movilidad a partir de la telefonía móvil, la medición del comercio diario al por menor de grandes empresas, el gasto realizado por los visitantes extranjeros en sus visitas a España, la medición del número de viviendas turísticas en España y su capacidad, o la medición del turismo nacional y receptor a partir de la posición de los teléfonos móviles. En este punto es necesario puntualizar que estas estadísticas experimentales no son exclusivas del ámbito nacional o europeo, sino que también han proliferado en otras zonas, por ejemplo, en Estados Unidos donde el Census Bureau ha puesto en marcha nuevas estadísticas como el *Census Household Pulse Survey* o el *Small Business Pulse Survey*.

A modo de ejemplo, la Figura 5 muestra la variación mensual de los nacimientos con respecto a 2019 del acumulado en lo que va de año, revelando un notable retroceso de esta variable en enero de 2021 con un descenso en el número de nacimientos del 22,84 %, que paulatinamente ha ido recuperándose hasta situarse en un –9,59 % en el mes de abril de 2021. Este gráfico también nos aporta el exceso de mortalidad en España, elaborado con periodicidad mensual y en comparación con la media mensual obtenida durante el periodo 2016-2019. El valor negativo obtenido en enero, febrero, junio de 2020 y marzo de 2021 nos indica un menor número de fallecidos durante esos meses comparados con el periodo base, en tanto que el máximo de la serie se obtiene en los meses de abril, marzo y noviembre de 2020.

FIGURA 5
VARIACIÓN DE LA NATALIDAD Y EL EXCESO DE MORTALIDAD EN 2020 Y 2021 POR MESES
(En %)



FUENTE: INE y elaboración propia.

Sin duda, las estadísticas experimentales anteriormente mencionadas amplían el sistema de información estadístico existente y han sido muy bien acogidas por los usuarios. No obstante esta medida también puede ser complementada por otras actuaciones que les permitan trabajar directamente con las bases de datos, para realizar investigaciones y estudios de interés público salvaguardando en todo caso el secreto estadístico, la confidencialidad y la protección de los datos. En este sentido, el reciente acuerdo alcanzado en abril de 2021 entre el INE, la Agencia Estatal de Administración Tributaria, la Seguridad Social, el Servicio Público de Empleo Estatal y el Banco de España permitirá incrementar la utilidad de la información administrativa disponible así como de los datos recogidos con fines estadísticos, habilitando el acceso de datos granulares a investigadores que desarrollen proyectos científicos de interés público. Se trata de un enfoque que da respuesta a muchas demandas de investigadores y organismos públicos, entre otros, al planteamiento realizado por la AIReF en su opinión para una estrategia de acceso a datos administrativos (AIReF, 2020b). No obstante, la puesta en marcha de un proyecto de este calibre requiere nuevas inversiones para actualizar las infraestructuras tecnológicas así como una adaptación del marco normativo existente, particularmente, la Ley de la Función Estadística Pública, para permitir un acceso a datos por parte de los investigadores, al menos, en las mismas condiciones al modelo ya establecido en el ámbito de la ley estadística europea.

Un caso evidente de creación de infraestructura tecnológica sería la implantación de un nodo de interoperabilidad que permita el intercambio de información entre los agentes implicados que se determinen. La información a intercambiar podría consistir en datos y ficheros procedentes de encuestas, de registros administrativos o de otras fuentes alternativas, en tanto que el conjunto de agentes implicados podrían ser desde los servicios estadísticos (estatales o autonómicos), servicios públicos no estadísticos, investigadores o entidades de investigación de diferente naturaleza jurídica. En este punto es necesario subrayar que los usos que se podrían hacer de dicha información deberían ser diferentes en función de las necesidades de los distintos y potenciales agentes mencionados anteriormente. Y, en todo caso, cualquier acceso y posible uso tendría que estar regulado por la legislación y ser objeto de la correspondiente monitorización. La disponibilidad de un sistema de integración y acceso a datos, tanto estadísticos como administrativos, que permita el acceso a los mismos de forma eficiente y con las condiciones de seguridad convenientes, nos permitiría alinearnos con las iniciativas más punteras que están surgiendo a nivel internacional para facilitar una mayor interoperabilidad de datos e información estadística.

Junto con la inversión tecnológica también resulta preciso adaptar los conocimientos del personal para poder afrontar una producción estadística basada en múltiples fuentes de información, favoreciendo la incorporación de profesionales con formación orientada a los métodos y herramientas para el análisis de datos masivos (estadísticos especializados en ciencia de datos) y, al mismo tiempo, de técnicos con amplios conocimientos de índole contable y demográfica (estadística aplicada a las ciencias económicas y sociales). Esta medida puede venir complementada con otras actuaciones para reforzar la colaboración con las universidades, promoviendo la realización de proyectos de investigación científica que permitan el desarrollo de nuevos conocimientos y mejoren la producción de datos y estadísticas contribuyendo, además, a la I+D+i en estadística en áreas clave como pueden ser: la codificación automática con técnicas de *machine learning*, el lenguaje natural e inteligencia artificial aplicados a la recogida de datos, la publicación y uso de datos abiertos enlazados (*linked open data*), o la aplicación de estimación en pequeñas áreas para el uso de información auxiliar en la desagregación de la información de encuestas, entre otros. Es necesario apuntar que este nuevo paradigma de datos e información estadística se encuentra esbozado en las propias líneas estratégicas del Programa anual 2021 del Plan Estadístico Nacional 2021-2024, aprobado por medio del Real Decreto 150/2021 de 9 de marzo, que entre otros incluye el incremento de la explotación de los registros administrativos en la producción de estadísticas oficiales; el establecimiento de una arquitectura de datos comunes procedentes de registros administrativos asociados a los principales directorios con un tratamiento armonizado; inclusión de identificadores estandarizados; el análisis de la creación de plataformas seguras para almacenar registros administrativos para su posterior explotación estadística; o la creación de un Sistema integrado de Información de las Empresas.

Para concluir, debemos tener presente los notables avances que la Unión Europea está realizando en materia de digitalización y, particularmente, el impulso que se está

otorgando tanto a la creación de un espacio común europeo de datos así como a la reutilización de información del sector público. En efecto, la estrategia de datos europea facilita un marco de gobernanza intersectorial para el acceso a los datos y su uso, además de la creación de espacios comunes de datos en sectores estratégicos y de interés público. En este contexto resulta conveniente que el sistema estadístico pueda desarrollar funciones de administrador de datos (*data stewardship*) y participe en el diseño de una estructura de gobernanza que facilite el intercambio de datos *Business to Government* (*B2G*). De hecho países como Canadá, Estonia, Irlanda, Reino Unido o Nueva Zelanda son punteros a la hora de buscar sinergias entre los roles de *Chief Statistician* y *Chief Data Steward*. Una situación en la que los servicios estadísticos puedan actuar como *data steward* ofrecería múltiples oportunidades para la producción estadística oficial, al permitir la utilización de nuevas fuentes de información en la producción estadística, reduciendo el coste y carga de las unidades informantes al combinar datos estadísticos con registros administrativos y otras fuentes de datos privadas, permitiendo a su vez mejorar la utilidad social de los datos, en particular para la evaluación de políticas públicas, que pueden beneficiarse de la contrastada experiencia del INE en el tratamiento y normalización de los datos. Por ello resulta conveniente avanzar en el diseño de una «estrategia estadística del dato» que consiga integrar las perspectivas *Government to Government* (*G2G*), *Government to Business* (*G2B*) y *Business to Government* (*B2G*) y fomente en particular el acceso a los datos para investigación (Consejo Superior de Estadística, 2021). En particular, el próximo *Data Act* que se aprobará en el seno de la Unión Europea proporcionará una oportunidad única para promover la compartición de datos *B2G* para fines de interés público, en especial para la estadística oficial (es decir, *B2G4S*) manteniendo en todo caso las condiciones de salvaguarda, confidencialidad y secreto estadístico de la información.

5. Conclusiones

Una estadística oficial de calidad constituye un elemento esencial para el proceso de toma de decisiones y la evaluación de políticas públicas. Durante décadas los servicios estadísticos han cumplido su misión con éxito utilizando información procedente de las fuentes de datos tradicionales (censos, encuestas y registros administrativos) si bien este modelo ya no resulta del todo suficiente. La senda de modernización de productos y procesos estadísticos ha estado principalmente orientada a incrementar tanto la elaboración de nuevas estadísticas como el grado de satisfacción de los usuarios. La crisis de la COVID-19 ha demostrado la resiliencia del INE y de la estadística oficial en su conjunto, que están sólidamente fundamentados, si bien aún quedan cuestiones por resolver particularmente en lo que respecta al uso del *big data* en la estadística oficial.

La utilización de estas nuevas fuentes de datos en la estadística oficial conlleva la necesidad de adaptar los enfoques de la producción estadística, las capacidades y

competencias de los estadísticos, los métodos de producción y las tecnologías utilizadas, buscando que nuestras estadísticas sigan siendo relevantes y significativas, y manteniendo en todo caso el alto grado de confianza que tiene la sociedad en los servicios estadísticos. Para ello es necesario poder acceder a las nuevas fuentes de datos e integrarlas junto con las otras fuentes de datos más tradicionales para la producción de estadísticas oficiales. Es necesario también disponer de recursos humanos y tecnológicos suficientes, incluyendo procesos, estándares y metodologías armonizadas a nivel europeo y, todo ello, dentro de un marco de preservación de la privacidad y con la máxima garantía en la seguridad de los datos.

Para superar exitosamente estos retos es necesario, en primer lugar, adaptar la legislación nacional existente. Las nuevas reglas deben tener como objetivo maximizar el beneficio para toda la sociedad equilibrando los derechos e intereses legítimos de todas las partes interesadas involucradas. El equilibrio de intereses y derechos debe basarse, por un lado, en el derecho de los ciudadanos a la privacidad individual como sujetos de datos y a vivir en una sociedad democrática informada como usuarios de estadísticas y, por otro lado, en el respeto a los intereses comerciales de las empresas que custodian los datos así como los intereses comerciales de todos los demás actores económicos para tomar sus decisiones informadas basadas en estadísticas públicas. Varias actuaciones se han puesto en marcha en paralelo para hacer frente a esta situación: a nivel europeo se están tomando iniciativas para que las necesidades estadísticas estén convenientemente recogidas en la futura Ley del Dato Europea (*Data Act*) y, en caso de ser necesario, en la ley estadística europea; y a nivel nacional el Plan Estadístico Nacional ya contempla como una de las líneas estratégicas para este periodo la reforma de nuestra ley estadística, la Ley de la Función Estadística Pública, en particular para recoger estos aspectos anteriormente señalados.

En segundo lugar, como ya se ha mencionado, es necesario abordar la modernización de la producción estadística, en especial las infraestructuras tecnológicas que favorezcan la interoperabilidad de los datos entre el conjunto de agentes que se determinen. Adicionalmente es necesario adecuar los conocimientos y capacidades del personal incluyendo una especialización y adaptación de las pruebas de acceso a los cuerpos de estadística del estado, potenciando nuevos cursos de formación, el uso de nuevas herramientas dirigidas a alcanzar este objetivo así como la colaboración con el mundo académico y universitario.

Por último, es importante que en todo momento quede garantizado el mismo nivel de armonización y alta calidad de las estadísticas basadas en *big data* en relación con el que gozan las estadísticas anteriormente calificadas como tradicionales. Para ello se necesita seguir participando en los estudios piloto a nivel europeo en los que se testan nuevas metodologías, a la vez que continuar avanzando en esta línea con las estadísticas experimentales, y buscar la colaboración de los medios académicos y del sector privado, para avanzar conjuntamente y alcanzar el objetivo común que no es otro que la mejora de nuestra sociedad.

Referencias bibliográficas

AIReF, Autoridad Independiente de Responsabilidad Fiscal (2020a). *Estudio de Beneficios Fiscales de la Fase II del Spending Review*. https://www.airef.es/wp-content/uploads/2020/10/Docus_Varios_SR/Estudio_Beneficios_Fiscales_Spending_Review.pdf

AIReF, Autoridad Independiente de Responsabilidad Fiscal (2020b). *Opinión para una estrategia de acceso a datos administrativos*. https://www.airef.es/wp-content/uploads/2020/10/OPINION_DATOS/201005.-Opinion-Acceso-Datos.pdf

Arnaboldi, M., & Azzone, G. (2020). Data science in the design of public policies: Dispelling the obscurity in matching policy demand and data offer. *Heliyon*, *6*(6), e04300. https://doi.org/10.1016/j.heliyon.2020.e04300

Azzone, G. (2018). Big data and public policies: Opportunities and challenges. *Statistics & Probability Letters*, *136*, 116-120. https://doi.org/10.1016/j.spl.2018.02.022

Barreto-Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de población*, *18*(73), 241-271.

Card, D., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *American Economic Association, ten years and beyond: Economists answer NSF's call for long-term research agendas*. National Science Foundation

CE (1997). Reglamento (CE) nº 322/97 del Consejo de 17 de febrero de 1997 sobre la estadística comunitaria. *Diario Oficial de las Comunidades Europeas*. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:31997R0322

CE (2009). Reglamento (CE) nº 223/2009 del Parlamento Europeo y del Consejo de 11 de marzo de 2009 relativo a la estadística europea. *Diario Oficial de la Unión Europea*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02009R0223-20150608

Consejo Superior de Estadística (2018). *Propuestas y recomendaciones, previas a la formulación del anteproyecto del Plan Estadístico Nacional 2021-2024, sobre las necesidades nacionales en materia estadística y la adaptación y mejora de los medios existentes*. https://www.ine.es/normativa/leyes/plan/plan_2021-2024/recomendaciones_2021-2024_cse.pdf

Consejo Superior de Estadística (2021). *El papel de la estadística oficial en la administración y gestión de datos*. https://www.ine.es/normativa/leyes/cse/papel_estadistica_oficial.pdf

Crato, N., & Paruolo, P. (2019). Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design. *Springer Open*. https://doi.org/10.1007/978-3-319-78461-8

ECA, European Court of Auditors. (2021, 4 de mayo). *Los auditores estudian la calidad de las estadísticas de la UE*. [comunicado de prensa] https://www.eca.europa.eu/Lists/ECADocuments/INAP21_05/INAP_High-quality-EU-stats_ES.pdf

Eurostat (2012). *Quality Assurance Framework of the European Statistical System*. https://ec.europa.eu/eurostat/documents/3859598/5923349/QAF_2012-EN.PDF.pdf/fcdf3c44-8ab8-41b8-9fd0-91bd1299e3ef?t=1414781890000

Eurostat (2014). *Towards a harmonised methodology for statistical indicators.Part 1, Indicator typologies and terminologies: 2014 edition*. https://doi.org/10.2785/56118

Eurostat (2018). *Código de buenas prácticas de las estadísticas europeas*. https://doi.org/10.2785/232692

Eurostat (2020). *European Statistical System (ESS) handbook for quality and metadata reports*. https://doi.org/10.2785/666412

Fano, M. P. (2017). Un nuevo marco legal para la evaluación. *Gestión y análisis de políticas públicas*, (17), 6-25. https://doi.org/10.24965/gapp.v0i17.10397

Feinstein, O. (2007). Evaluación pragmática de políticas públicas. *ICE, Revista de Economía* (836).

Ley 12/1989, de 9 de mayo, de la Función Estadística Pública. *Boletín Oficial del Estado*, nº 112, de 11 de mayo de 1989, pp. 14026 a 14035. https://www.boe.es/eli/es/l/1989/05/09/12

Listorti, G., Basyte-Ferrari, E., Acs, S., & Smits, P. (2020). Towards an Evidence-Based and Integrated Policy Cycle in the EU: A Review of the Debate on the Better Regulation Agenda. *JCMS: Journal of Common Market Studies*, *58*(6), 1558-1577. https://doi.org/10.1111/jcms.13053

Loi, M., & Rodrigues, M. (2012). *A note on the impact evaluation of public policies: The counterfactual analysis*. JRC. https://doi.org/10.2788/50327

Merediz, A. (2004). *Historia de la estadística oficial como institución pública en España*. Instituto de Estadística de Andalucía. https://www.juntadeandalucia.es/institutodeestadisticaycartografia/historicas/HistEstOficial/HistEstOficial.pdf

Moral, I., Paniagua, M., Rodríguez, L., & Rodríguez, C. (2016). *Evaluación de políticas públicas. Técnicas cuantitativas*. Garceta

OCDE, Organización para la Cooperación y el Desarrollo Económico. (2020a). *How's Life? Measuring Well-being*. https://doi.org/10.1787/9870c393-en

OCDE, Organización para la Cooperación y el Desarrollo Económico. (2020b). *Improving governance with policy evaluation: Lessons from country experiences*. https://doi.org/10.1787/89b1577d-en

Pollitt, C., & Bouckaert, G. (2010). *La reforma de la gestión pública: Un análisis comparado*. INAP.

Radermacher, W. J. (2019). Governing-by-the-numbers/Statistical governance: Reflections on the future of official statistics in a digital and globalised society. *Statistical Journal of the IAOS*, *35*(4), 519-537. https://doi.org/10.3233/SJI-190562

Rico, J. (2020). *Guía de evaluación de resultados de políticas públicas*. Instituto para la Evaluación de Políticas Públicas. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia_de_Evaluacion_de_Resultados.pdf#page=1

Statistics Canada (2003). *Survey methods and practices*.

Tratado de Ámsterdam por el que se modifica el tratado de la Unión Europea, los tratados constitutivos de las Comunidades Europeas y determinados actos conexos. *Diario Oficial de las Comunidades Europeas*. C340, de 10 de noviembre de 1997. http://data.europa.eu/eli/treaty/ams/sign

UNECE, United Nations Economic Commission for Europe. (2009). *Cómo hacer comprensibles los datos. Parte 2: Una guía para presentar estadísticas*. https://unece.org/DAM/stats/documents/writing/MDM_Part2_Spanish.pdf

UNECE, United Nations Economic Commission for Europe. (2018). *Guidance on Modernizing Statistical Legislation*. https://unece.org/DAM/stats/publications/2018/ECECESSTAT20183.pdf

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Investigación económica y datos masivos: mercados, fines sociales y colaboración público-privada*

Economic research and big data: Markets, socio-economic research and public-private partnerships

José García Montalvo
Universitat Pompeu Fabra

Resumen

Este artículo presenta un análisis de las perspectivas abiertas por la creciente disponibilidad de datos masivos para realizar investigación económica, así como los riesgos asociados a las mismas. A diferencia de otras contribuciones, la aproximación adoptada se enfoca en los generadores de la información distinguiendo entre los datos generados por empresas privadas para buscar soluciones de mercado, los datos originados en las administraciones públicas y las experiencias recientes de colaboración público-privada que se están abriendo en el campo del uso de datos masivos (big data) y la aplicación de técnicas de aprendizaje automático (machine learning).

Palabras clave: *datos masivos, aprendizaje automático, colaboración público-privada.*
Códigos JEL: *C10, C18, C55, C81.*

Abstract

This article presents an analysis of the opportunities opened by the increasing availability of big data for economic research as well as the risks associated with them. Unlike other contributions, the approach adopted here focuses on the generators of the information, distinguishing between data generated by private companies to seek market solutions, the data originating in public administrations used for research, and the recent experiences of public-private partnerships that are opening large datasets of private companies for research in empirical economics using big data and machine learning techniques.

Keywords: *big data, machine learning, public-private partnership.*

* Este artículo se basa fundamentalmente en una conferencia impartida en la Fundación Ramón Areces titulada «*Big data* y economía en la era de las *fake news* y las verdades imprecisas».

DOI: https://doi.org/10.32796/cice.2021.102.7311
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

Este artículo presenta un recorrido por las oportunidades abiertas por la creciente disponibilidad de datos masivos para realizar investigación económica y los riesgos asociados a las mismas. La aproximación adoptada distingue entre los datos generados por empresas privadas para buscar soluciones de mercado, los datos originados en las administraciones públicas y las experiencias recientes de colaboración público-privada que se están abriendo en el campo del uso de datos masivos (*big data*) y la aplicación de técnicas de aprendizaje automático (*machine learning*). Por datos masivos entendemos tanto lo que normalmente se describe como *big data* (datos no tradicionales con formato muy heterogéneo como sensores, GPS, *clicks*, *logs* de servidores, correos electrónicos, imágenes, voz, texto de redes sociales, etc.) así como datos estructurados de fuentes administrativas o empresas. Este artículo es, en parte, continuación de García-Montalvo (2014) donde se realiza una panorámica de la utilización de datos masivos con objetivos predictivos en el sistema bancario y financiero. En el mismo se describen sistemas de *scoring* basados en información de las redes sociales, y procedimientos innovadores de calificación crediticia fundamentados en el análisis textual de las solicitudes digitales. Asimismo, se describen procedimientos de detección de fraude en tarjetas de crédito, de personalización de las tarifas de las compañías de seguros, etc. García-Montalvo (2021) describe un conjunto amplio de experiencias personales en el trabajo de investigación económica utilizando grandes bases de datos administrativos y técnicas de aprendizaje automático.

2. Datos privados para soluciones de mercado

El dominio de las bases de datos masivas comienza en los años noventa con la mejora de los sistemas de información y el aumento de la potencia de los ordenadores. Las primeras aplicaciones de datos masivos se producen en las grandes infraestructuras científicas como los sincrotrones. De hecho, aunque se piensa que en la actualidad es Internet el mayor generador de datos, esta idea es errónea. El Large Hadron Collider del CERN produce 600 TB/segundo con sus 15 millones de sensores y, después de un filtrado que desecha la mayor parte de la información, necesita almacenar 25 PB/año. De hecho, el problema fundamental en la actualidad para el avance del trabajo con bases de datos cada vez más grandes no es la capacidad de cálculo, sino que la creación de información avanza a más velocidad que la capacidad de almacenarla y la energía necesaria para mover dicha información entre los procesadores y los dispositivos de almacenaje es cada vez mayor.

La mayoría de las primeras aplicaciones basadas en datos masivos en el campo de la economía y las finanzas tienen que ver con datos generados en mercados financieros. Es la época del surgimiento de los *quants*[1]. En general, los objetivos son poner precio a productos financieros complejos, analizar correlaciones entre

[1] Para una descripción muy interesante de esta época, consúltese Derman (2004).

enormes conjuntos de activos financieros o buscar estrategias para aprovechar pequeñas disfuncionalidades de precios que se producían entre mercados o activos. Pero también se podían utilizar para hacer evaluaciones de la implementación de medidas para mejorar la liquidez de un mercado. Por ejemplo, García Montalvo (2003) analiza, usando decenas de millones de observaciones de *bids*, *asks* y operaciones cerradas en el mercado de futuros de deuda pública, el impacto de la introducción de creadores de mercado en 1995 en el Mercado Oficial de Futuros y Opciones Financieras de España (MEFF) sobre la liquidez del mercado.

La llegada de procedimientos de aprendizaje automático más avanzados[2], favorecidos por la creciente potencia de los ordenadores, facilitó la extensión de la utilización de los datos para la solución de problemas empresariales. La utilización de datos masivos para la solución de problemas de mercado viene marcada por varios hitos. La sustitución de los críticos y editores de Amazon por un sistema automático de recomendación basado en el algoritmo *item-by-item* fue uno de los primeros grandes logros de la recopilación masiva de información sobre compras de los clientes de Amazon. Pero el primer gran hito es el caso Target. Imaginémonos a principios de los años 2000. Un coche recorre las nevadas llanuras de Minnesota hasta llegar a un centro comercial donde para enfrente de una tienda de la cadena Target. Un hombre sale del coche y se dirige al interior donde solicita hablar con el encargado. Cuando este llega, el cliente desparrama sobre el mostrador un montón de vales con ofertas de pañales de niños, carritos de bebé, muñecos, potitos, etc. El cliente, muy enfadado, grita: «¡Cómo se atreven a enviarle estos vales a mi hija! ¡Solo tiene 15 años!». El encargado intenta tranquilizar al cliente y le pide disculpas argumentando que a veces los sistemas fallan y que siente el error. El cliente sale de la tienda, todavía muy enfadado, y sube al coche desapareciendo en el horizonte.

A los pocos días, el encargado telefonea al cliente para volver a disculparse por el error. Al sonar el teléfono e identificarse, el cliente detiene su intento de disculpa. «Quizá quien tendría que disculparse soy yo. Bajo el techo de mi casa estaban sucediendo cosas que yo mismo desconocía.» En efecto, su hija adolescente estaba embarazada y Target lo supo antes que su propio padre[3]. El procedimiento utilizaba la combinación de productos en una cesta de la compra para elaborar tipologías de clientes. Las mujeres embarazadas son un objetivo frecuente de los modelos predictivos, puesto que la captación en esta etapa garantiza compras futuras y fidelidad temporal. Este tipo de predicción fue posible por la enorme cantidad de información disponible a partir de la utilización de los lectores de códigos de barras y las mejoras en los métodos de aprendizaje automático.

---

[2] La generalización de procedimientos como el *support vector machine,* las técnicas de *random forest* y el redescubrimiento de las redes neuronales que, después de las limitaciones mostradas en los años sesenta, ahora pasan a denominarse *deep learning* para intentar evitar el estigma asociado a estas técnicas, se ve favorecida por la creciente capacidad de cálculo de los ordenadores.

[3] Esta historia se conoce porque Andrew Pole, el estadístico a cargo de los primeros modelos predictivos de Target, la contó en una conferencia muchos años después, y el *New York Times* se hizo eco, apareciendo también en Duhigg (2011).

El segundo hito importante es el premio de Netflix. La competición premiaba con un millón de dólares al mejor algoritmo colaborativo para predecir los *ratings* de las películas de los clientes que superara en un 10 % la capacidad predictiva del modelo utilizado por Netflix en el momento. La competición comenzó en octubre de 2006. Netflix proporcionó unos 100 millones de *ratings* de unos 500.000 usuarios sobre 17.770 películas. La dificultad consistía en que el modelo tenía que predecir únicamente utilizando la información sobre los *ratings* pasados sin ninguna información sobre usuarios o películas. En septiembre de 2009, el grupo BellKor's Pragmatic Chaos triunfó en la competición, en la que participaron 2.000 equipos de 150 países, incluidos equipos universitarios de los mejores centros del mundo. La mejora del grupo sobre el algoritmo original fue del 10,06 %.

La multiplicación de la utilización de datos para solucionar problemas de mercado es interminable. En servicios financieros las nuevas Fintech utilizan procedimientos de calificación basados en datos de redes sociales. Por ejemplo, quiénes son tus amigos en Facebook y cuántos están en listas de morosos, quiénes son tus contactos en LinkedIn (número de jefes en la empresa actual o directivos de otras empresas), el análisis textual de solicitudes de crédito o incluso el hecho de capitalizar o no capitalizar el nombre y apellidos en una solicitud. En el campo de los seguros de auto la aceptación de sensores en el vehículo permite un perfilado muy preciso del nivel de riesgo de un conductor. Pero sin duda en el campo de la publicidad en internet es donde más se ha avanzado con subastas de posicionamientos casi instantáneas.

La aproximación a la solución de estos problemas también ha cambiado. Una forma de visualizar la evolución de estos procedimientos consiste en analizar como la inteligencia artificial ha ido destronando a los mejores jugadores de ajedrez y Go. En 1997, y después de algunos intentos, Deep Blue consiguió vencer a Kasparov zanjando de forma definitiva la disputa de los ajedrecistas con las máquinas. La siguiente frontera era un juego mucho más complejo: el Go. El AlphaGo de Google venció en 2015 a un segundo dan del Go por 5 a 0. Pero el momento definitivo llegó con la victoria sobre Lee Sedol, héroe nacional de Corea y mejor jugador de la historia. AlphaGo ganó a Sedol en el Google DeepMind Challenge de 2016 por 4 a 1. Lo importante no es la demostración del dominio de las máquinas en estos juegos sino la evolución de los procedimientos. Mientras en los sistemas de inteligencia artificial clásicos se introducían las reglas y miles de jugadas de partidas históricas, AlphaGo funciona de una forma muy diferente. Sus procedimientos de redes neuronales aprenden jugando millones de partidas… contra sí mismo. Este sistema permite acumular una cantidad ingente de jugadas, muchas de ellas nunca utilizadas por humanos, así como los resultados de las diferentes estrategias. Los procedimientos de aprendizaje reforzado consiguen evitar los vicios y servidumbres de las jugadas humanas del pasado y los programas anteriores.

3. Datos masivos para el avance social

Los datos masivos y los procedimientos de aprendizaje automático han pasado también a ser la base de una creciente cantidad de investigación social y económica. Un problema de los métodos tradicionales de recopilación de datos, como las encuestas, es que cada vez tienen un nivel de no respuesta superior (Meyer *et al*., 2015) y los encuestados mienten con más frecuencia, en particular en preguntas relacionadas con magnitudes económica importantes como la renta. Además, los datos de encuestas retrospectivas tienen problemas con la fidelidad de las respuestas dado que las preguntas retrospectivas suelen tener un alto nivel de error de medida. Card *et al*. (2010) describen los datos administrativos como muy superiores a las encuestas porque son muestras mucho más grandes (toda la población), son longitudinales por naturaleza y tienen menos problemas de medida, imputaciones o pérdida de muestra[4]. Finalmente, las operaciones estadísticas basadas en encuestas suponen disponer de la información con bastante retraso.

Los datos administrativos y algunos de los datos obtenidos en Internet no presentan estos problemas[5]. Por ejemplo, Davidowitz (2017) define las búsquedas en Internet como el suero de la verdad. Todo el mundo miente… excepto cuando busca en Internet[6]. Por ejemplo, cuando se escribe en el buscador «es mi marido» el buscador completa con el término «homosexual» (al menos en el momento que Davidowitz escribió el libro) en lugar de «infiel» que sería quizás lo primero que nos vendría a la cabeza. Uno podría pensar que existe una gran correlación entre el desempleo en Estados Unidos y las búsquedas de «oficina de empleo» o «vacante». Sin embargo, en momentos de mayor desempleo crecen las búsquedas de «porno» y de juegos de solitario. Se busca con más frecuencia «tamaño de pene» o «alargamiento de pene» que como cambiar una rueda. Se busca mucho más «mi hijo es superdotado» que «mi hija es superdotada» y, sin embargo, se busca más «mi hija es obesa/guapa/fea» que «mi hijo» cuando los niños tienen estadísticamente una probabilidad superior de ser obesos que las niñas.

La disponibilidad creciente de enormes bases de datos, en muchos casos geocodificadas, de procedencia muy diversa hace de la economía una disciplina cada vez más científica. Un ejemplo destacado de este tipo de estudios es el Billion Prices Project (Cavallo & Rigobon, 2016), que estima la evolución en tiempo real de la evolución de los precios utilizando millones de precios de tiendas *online*. Más allá de los resultados estrictamente científicos, el proyecto permitió comparar la evolución

[4] Esta visión contrasta con la aproximación en Barcellan *et al*. (2017) donde el mensaje básico es que no existe ningún tipo de dato superior a otro.

[5] En este sentido, las redes sociales son menos útiles. Por ejemplo, multitud de investigadores han intentado utilizar análisis del sentimiento en Twitter y Facebook para predecir el resultado de elecciones con muy poco éxito. Han tenido más éxito las aplicaciones del análisis textual a las actas de comités de los bancos centrales (Hansen *et al*., 2018).

[6] Este es el motivo por el que Google Trends es una herramienta cada vez más utilizada para hacer todo tipo de predicciones.

de la inflación oficial y de los precios *online* en diversos países de Latinoamérica. Con esta información se pudo comprobar que los patrones eran similares en Brasil, Colombia o Chile pero no en Argentina, donde la diferencia acumulada de la inflación entre 2007 y 2011 fue un 65 % (Cavallo, 2013).

Otros datos que se utilizan cada vez más intensivamente son las imágenes de satélites. Diversos estudios han comprobado la elevada correlación de la luz nocturna con el nivel de desarrollo (Chen & Nordhaus, 2011), la bondad relativa del PIB per cápita y la renta media de las encuestas de presupuestos familiares como medida del desarrollo económico (Pinkovskiy & Sala-i-Martin, 2016) o la desigualdad (García Montalvo *et al*., 2021). Más recientemente, también se está avanzando en complementar la medida a partir de la luz nocturna con imágenes de día en línea con las propuestas originales en Jean *et al*. (2016). Además, los datos de luz nocturna se usan rutinariamente para medir el nivel de desarrollo en áreas pequeñas que no coinciden con unidades administrativas y no tienen información sobre renta o riqueza (García Montalvo & Reynal-Querol, 2021).

Otra fuente de datos que genera una cantidad ingente de información son los teléfonos móviles. Durante el periodo epidémico la geolocalización de los móviles ha servido para construir índices de movilidad para analizar el impacto de la COVID-19 en el movimiento de las personas (Chen *et al*., 2020). Pero quizás una de las propuestas más innovadora sea la que se presenta en Athey *et al*. (2020). Usando datos de GPS obtenidos a partir de teléfonos móviles Athey *et al*. (2020) analiza una nueva medida de segregación denominada aislamiento experimentado, que analiza la exposición de individuos de distintas razas, nivel socioeconómico, etc. en los distintos sitios que visitan a lo largo del día. Con esta información muestran que el aislamiento experimentado es menor que la segregación residencial, aunque ambas medidas están muy correlacionadas por ciudades y áreas geográficas.

Finalmente, otra fuente reciente de datos administrativos masivos muy útiles para la economía es la información de las agencias tributarias. Estos organismos han sido generalmente muy refractarios a compartir cualquier tipo de información, pero en fechas recientes la situación está cambiando. Probablemente este cambio esté en parte relacionado con el éxito del programa de investigación de Chetty y coautores sobre el análisis de la movilidad social y la igualdad de oportunidades en Estados Unidos[7]. Aunque inicialmente hubo cierta polémica sobre si la utilización de los datos se ajustaba a la solicitud original, los espectaculares resultados obtenidos acallaron estas críticas. A partir de esta información es posible analizar en áreas geográficas muy pequeñas, la relación entre la posición en la distribución de la renta de los hijos y sus padres. Uno de los resultados más impactantes se refiere a la proporción de hijos que ganan más que ganaban sus padres en Estados Unidos: mientras en 1940 este porcentaje alcanzaba el 92 %, en 1984 era solo el 50 %, lo que muestra una caída espectacular (Chetty *et al*., 2016).

[7] https://opportunityinsights.org/

El dominio creciente de los datos está cambiando incluso la forma de explicar la economía en los cursos básicos. La puesta en marcha del proyecto CORE *(Curriculum Open-access Resources in Economics)* en 2013 supuso un cambio significativo en cómo se enseñaba economía en muchas universidades. No obstante, en muchas de las grandes universidades los cursos introductorios seguían guiándose por la visión clásica de la enseñanza de los principios de economía. Pero la enseñanza de los principios básicos de la economía basados únicamente en análisis empíricos está llegando a los cursos introductorios. Un ejemplo sintomático es el curso EC10 de la Universidad de Harvard. Este es el curso básico de principios de economía que realizan muchos universitarios en Harvard, desde biólogos hasta físicos y, por supuesto, estudiantes que quieren hacer el *major* en economía. Hasta el año 2019 el curso tuvo una estructura habitual, con muchas curvas de indiferencia, restricciones presupuestarias y cálculo elemental. La sustitución de Greg Mankiw por Raj Chetty como instructor del curso supuso una revolución absoluta en la forma de enseñar el curso. Las curvas de indiferencia dejaron paso a los datos y las cuestiones relativas a la distribución de la renta incrementaron su presencia sustancialmente.

4. La colaboración público-privada

Tradicionalmente la mayor parte de los datos se generaban y almacenaban en instituciones públicas (institutos de estadística, organismos oficiales, etc.) donde el acceso estaba claramente definido[8]. En la actualidad la mayor parte de la generación de datos se encuentra en el sector privado. De esta forma es necesario establecer colaboraciones con las empresas generadoras de datos para poder aprovechar su potencial en la evaluación de políticas públicas, el seguimiento de la economía, etc.

El advenimiento de la pandemia de COVID-19 y la necesidad de contar con información de muy alta frecuencia para hacer un seguimiento de la evolución económica y del impacto de las políticas adoptadas para mitigar el efecto de la pandemia ha hecho más importante la disponibilidad rápida de grandes bases de datos, impulsando la colaboración público-privada. En este caso el calificativo público puede referirse a universidades, organismos públicos y centros de investigación trabajando con grandes bases de datos de empresas privadas, aunque también se puede referir a bases de datos públicas construidas a partir de microdatos de empresas privadas.

Un ejemplo de esta colaboración es el *economic tracker* del *Opportunity Insight*. Chetty *et al*. (2020) describen los datos que utilizan para hacer un seguimiento del impacto de la COVID-19 en diferentes dimensiones. Entre las empresas privadas que colaboran aportando datos están Affinity Solutions (gasto en tarjetas de crédito y débito), CoinOut (transacciones en efectivo), Womply (utilización de tarjetas de crédito en pequeños negocios), Paychex (datos sobre empleo y nóminas de 670.000 pequeñas y medianas empresas), Intuit (ofrece servicios de pago de nóminas a

[8] Otro problema era la cuestión de la coordinación entre diferentes instituciones, los requerimientos para su utilización, etc. Estos temas serán comentados con posterioridad.

empresas), Earnin (ofrece servicios de avance de la nómina que permite acceder a la renta de los trabajadores que se han inscrito en la web), Kronos (servicio de gestión de trabajadores que comprende a 30.000 empresas y 3.2 millones de trabajadores), Homebase (herramienta de gestión de pequeñas empresas) y Zearn (aplicación de matemáticas utilizada por 925.000 estudiantes).

Otro ejemplo de colaboración público-privada son los acuerdos entre universidades y multitud de bancos para utilizar la información de las transacciones en sus cuentas. El listado en estos momentos incluye ya muchos países: Estados Unidos, Reino Unido, Francia, Japón, Dinamarca, Suecia, Islandia, Portugal, Países Bajos, China y España. Las empresas colaboradoras incluyen todo tipo de entidades financieras desde Fintechs hasta bancos tradicionales. El Cuadro 1 recoge un resumen de las investigaciones en marcha y las entidades privadas que les dan soporte.

CUADRO 1
EQUIPOS DE INVESTIGACIÓN Y ENTIDADES FINANCIERAS COLABORADORAS

Autores	País	Entidad privada	Usuarios
Agregadores financieros (aplicaciones financieras y Fintech)			
Kuchler *et al*. (2018)	Estados Unidos	Ready to Zero	516
Hacioglu *et al*. (2021)	Reino Unido	Money Dashboard	8.350
Baker *et al*. (2020)	Estados Unidos	SaverLife	5.746
Baker (2018)	Estados Unidos	Confidencial	156.606
Olaffson & Pagel (2018)	Islandia	Meniga	66.262
Olaffson *et al*. (2018)	Islandia	Meniga	55.545
G. Montalvo & Reynal (2020)	España	Fintonic	236.053
Gelman *et al*. (2014)	Estados Unidos	Check	23.000
Bancos tradicionales			
Bounie *et al*. (2021)	Francia	CIC (Credit Mutuel)/ CB (Carte bancaire)	300.000/1,8 millones de afiliados a CB
Sheridan *et al*. (2020)	Dinamarca/Suecia	Danske Bank	860.000
Aspachs *et al*. (2020)	España	Caixabank	3.028.204
Cox *et al*. (2020)	Estados Unidos	JP Morgan	5.014.672
Carvalho *et al*. (2021)	España	BBVA (tarjetas y TPV)	2.200.000 empresas
Kubota *et al*. (2021)	Japón	Mizuho Bank	2.800.000
Golec *et al*. (2020)	Países Bajos	ABN AMRO	2.000.000
Carvalho *et al*. (2020)	Portugal	Sociedade Interbancária de Serviços	

FUENTE: Elaboración propia.

El objetivo de estas colaboraciones es analizar teorías económicas como la renta permanente (Gelman *et al*., 2014; Olaffson *et al*., 2018), el efecto de los confinamientos (Chetty *et al*., 2020; Sheridan *et al*., 2020), el impacto de las políticas de transferencias de rentas frente a la COVID-19 (Kubota *et al*., 2021; Baker *et al*., 2020), la evolución del gasto durante la pandemia (Carvalho *et al*., 2021; Hacioglu *et al*., 2021; Bouni, 2021), el impacto distribucional de la epidemia sobre el gasto (García Montalvo & Reynal-Querol 2020; Hacioglu *et al*., 2021) o la evolución de la desigualdad mensual (Aspachs *et al*., 2021).

En el caso de España también se están produciendo ejemplos de colaboración entre entidades financieras y equipos de investigación. Un ejemplo es la colaboración entre un equipo de investigadores del departamento de Economía de la Universidad Pompeu Fabra y la unidad de investigación de Caixabank en el análisis en tiempo real de la desigualdad en España desde el comienzo de la epidemia de COVID-19 (Aspach *et al*., 2021)[9] o la colaboración de varios investigadores con el BBVA en la elaboración de información de muy alta frecuencia sobre la evolución del gasto en tarjeta de crédito por tipología de gasto (Carvalho *et al*., 2021). Asimismo, García Montalvo y Reynal-Querol (2020) utilizan los datos de la aplicación de agregación financiera Fintonic para analizar el efecto distribucional de la epidemia sobre el gasto. En todos estos casos se trata de bases de datos enormes. Por ejemplo, Carvalho *et al*. (2021) utilizan 1.300 millones de operaciones con tarjetas y terminales del BBVA. García Montalvo y Reynal-Querol (2020) usan 350 millones de transacciones de la Fintech española Fintonic.

Otro ejemplo de uso de datos privados se produce en el caso de la colaboración entre instituciones públicas u organismos oficiales y empresas. Por ejemplo, el Banco Central Europeo (BCE) usa la base de datos privada AnaCredit (*analytical credit data sets*), que contiene datos armonizados de la zona euro en una base de datos única, para sus análisis prudenciales y supervisorios. Asimismo, el Bank for International Settlements (BIS) recolecta y procesa información bancaria confidencial en colaboración con bancos centrales y autoridades nacionales para su International Data Hub (Doerr *et al*., 2021).

No obstante, sorprende que algunas de estas colaboraciones reciban críticas generalizadas a pesar de perseguir fines sociales. Un ejemplo claro en el caso español fue el anuncio del INE de comprar datos de las compañías telefónicas sobre geolocalización de los clientes para estudiar la movilidad urbana. Cuando se anunció el estudio los medios de comunicación titularon de forma muy crítica: «Así va a rastrear el INE tu móvil durante 8 días: a qué operadoras afecta y cómo evitarlo», «Las operadoras cobrarán medio millón de euros por ceder los datos al INE», «El INE va a espiar tu móvil 8 días: ¿Qué puedes hacer?». Esta reacción del público sorprende sobre todo cuando la mayoría de los usuarios de un teléfono móvil están dispuestos a ceder toda su información personal dando acceso a todos los datos de su móvil para instalarse una simple aplicación de linterna. Sin embargo, si una institución pública hace un

[9] https://inequality-tracker.caixabankresearch.com/

estudio con datos de usuarios de telefonía totalmente anonimizados y agregados por zonas, ponen el grito en el cielo. En este sentido falta claramente una mayor sensibilización de la población sobre la importancia de las bases de datos masivos para la consecución de fines sociales.

5. Limitaciones de los datos masivos y las técnicas de aprendizaje automático

En las secciones anteriores se ha mostrado el interés que tienen los datos masivos y las técnicas de aprendizaje automático para la solución de problemas de mercado, la investigación con objetivo social y la evaluación de políticas públicas. En particular, disponer de grandes bases de datos permite realizar análisis con un nivel de granularidad geográfica muy grande y, en muchos casos, con poco retraso temporal. No obstante, también tiene limitaciones y peligros (García Montalvo, 2014). Evidentemente, el primer peligro es injerencia en la privacidad de las personas con el objetivo de mejorar la capacidad predictiva de los modelos o su utilización para objetivos poco éticos[10]. Este problema es particularmente relevante respecto a los datos producidos por empresas privadas.

En segundo lugar, los algoritmos suelen reproducir los sesgos y la discriminación existente en la realidad *(lack of algorithm fairness)*. Por ejemplo, un algoritmo de *scoring* crediticio puede discriminar a las mujeres o las personas afroamericanas sin necesidad de incluir el identificador del grupo como una de las variables clasificatorias puesto que dicha característica está muy correlacionada con otras de las docenas de variables que se incluyen en el modelo. Un ejemplo se produjo con la investigación que el Departamento de Servicios Financieros del Estado de Nueva York a las tarjetas de crédito de Apple porque el algoritmo cargaba tipos de interés más altos a las mujeres[11]. Además, los procedimientos de aprendizaje automático suelen ser cajas negras que generan correlaciones y buenas predicciones, pero no explican causalidades. En este sentido es cada vez más necesario hacer auditorias de los algoritmos para entender y evitar estos sesgos[12].

En tercer lugar, la calidad de los datos es también un desafío importante. Es bien conocido que una característica habitual de las bases de datos masivas es la baja proporción señal/ruido. Por tanto, precisan de grandes esfuerzos para la limpieza de los datos. Un ejemplo claro son los datos obtenidos en las redes sociales. También hay problemas de representatividad, muestreo y fusión de datos de diferentes fuentes. La posibilidad de que los errores en la captura, fusión o limpieza de los datos

[10] Un ejemplo bien conocido de este problema es el caso de Cambridge Analytics.

[11] El departamento de servicios financieros justificaba su actuación señalando que «cualquier algoritmo que, intencionadamente o no, resulte en un tratamiento discriminatorio de la mujer viola las leyes de Nueva York».

[12] De hecho uno de los motivos tradicionales para utilizar procedimientos de aprendizaje automático es evitar las docenas de sesgos en las decisiones humanas que han sido reconocidos por la economía comportamental y la psicología.

generen consecuencias negativas para los ciudadanos a partir de la aplicación de técnicas de aprendizaje automático a datos de baja calidad es real. Por ejemplo, la industria de generación de calificaciones crediticias a partir de datos capturados en Internet puede provocar serios perjuicios a solicitantes de créditos si sus datos tienen un elevado grado de error. Yu *et al*. (2014) analizaron la información disponible por varias agencias de generación de calificaciones crediticias a partir de datos de Internet. Yu *et al*. (2014) seleccionaron a cinco compañías de *big data* y obtuvo los informes sobre 15 voluntarios para el estudio. Los informes recibidos tenían multitud de errores, estimaciones desmesuradas (salario doble del real del solicitante), direcciones incorrectas, multitud de información faltante (incluidas cuentas en redes sociales, etc.). Ciertamente los datos de las empresas tradicionales de generación de *credit score* de consumidores también son muy mejorables. Un estudio de 2013 de la Federal Trade Commission de Estados Unidos señalaba que el 20 % de los informes crediticios de estas compañías contienen errores y un 5 % de estos errores resultaron en una rebaja del *credit score* que impidió a los clientes conseguir un crédito o les supuso pagar un tipo superior. El problema de las agencias que se basan en *big data* es que los clientes no tienen forma de saber cómo se ha calculado su *credit score*. No hay forma de confirmar independientemente la capacidad predictiva del algoritmo utilizado. De esta forma el consumidor puede acabar siendo afectado negativamente por un *credit score* calculado a partir de datos erróneos aunque, a diferencia del caso de las agencias tradicionales, es más difícil estimar si esto ha afectado a su capacidad de conseguir un crédito o el tipo[13].

Además, incluso si los datos y los procedimientos de predicción son adecuados, es preciso hacer un seguimiento constante de la capacidad predictiva de los modelos pues la captura de datos puede cambiar en el tiempo y perder capacidad explicativa. Un buen ejemplo es lo sucedido con Google Flu trends. Como se ha comentado anteriormente las búsquedas en Google son muy útiles en modelos predictivos. Ginsberg *et al*. (2009) propusieron usar las búsquedas de una bolsa de palabras relacionadas con los síntomas de la gripe para hacer un seguimiento en tiempo real de la evolución de la gripe estacional con una enorme precisión geográfica. De hecho, el Centro de Control de Enfermedades Infecciosas (CDC) de Estados Unidos tardaba casi dos semanas en producir información sobre la evolución de la gripe y usando las búsquedas se podría proporcionar un indicador casi en tiempo real. La capacidad predictiva del procedimiento fue sorprendente en el periodo 2004-2011. Sin embargo, en 2012 la capacidad predictiva se deterioró enormemente como muestran Lazer *et al*. (2014). El motivo aparente de este problema es el cambio en el buscador de Google que sugiere palabras cuando se realiza una búsqueda.

[13] Garcia Montalvo (2014).

6. Investigación económica y acceso a datos masivos

Hace ya muchos años que, ante la previsible expansión de la utilización de bases de datos masivas, economistas académicos han analizado la necesidad de establecer procedimientos estandarizados para facilitar el trabajo de los investigadores. Card *et al*. (2010) presentan una propuesta para expandir el acceso a datos administrativos para investigación económica en Estados Unidos ante lo que describían como «erosión de la ventaja de EE. UU. en la creación de datos para la investigación económica». En aquel momento Card *et al*. (2010) ya se referían a los países escandinavos como ejemplo en la construcción, fusión y facilidad de acceso a grandes bases de datos administrativos. El llamado modelo nórdico ha sido desde siempre el espejo en el que muchos países han querido reflejarse a la hora de conseguir expandir el uso de datos administrativos para el uso en investigación. Tanto las oficinas estadísticas de Finlandia, Dinamarca, Noruega y Suecia permiten el acceso a los investigadores a grandes bases de datos administrativos ya fusionados. El acceso en el caso noruego también puede estar intermediado por el Norwegian Social Sciences Data Service. Más recientemente países como Alemania (IAB) y Francia (CASD) también han puesto en marcha organismos para facilitar el acceso a grandes bases de datos administrativos. Pero sin duda uno de los casos de éxito de los últimos años es la iniciativa del Reino Unido. Uno de los motivos fundamentales de esta iniciativa fue, de forma parecida a lo sucedido en Estados Unidos, la sensación de que el Reino Unido estaba perdiendo el tren de la investigación económica de frontera por las limitaciones que tenían los investigadores para acceder a grandes bases de datos administrativos. En 2014 se puso en marcha la Administrative Data Research Network del Reino Unido (ADRN) como una colaboración entre universidades, organismos gubernamentales, agencias estadísticas e investigadores para facilitar el acceso a datos administrativos ya fusionados entre diversas fuentes[14]. La iniciativa ha dado enormes frutos. En este sentido genera envidia sana que unos meses después de empezar la pandemia de COVID-19 investigadores del Reino Unido fueran capaces de fusionar 17 millones de historiales electrónicos de salud y calcular las probabilidad de mortalidad por COVID-19 en función de decenas de factores de riesgo (Williamson *et al*., 2020).

La situación en España, al menos hasta hace pocos años, fue de considerable retraso respecto a la situación de los países más avanzados en modelos de acceso a datos administrativos masivos. Esta situación de retraso es paradójica dado que los organismos públicos en España generan gran cantidad de datos administrativos de excelente calidad. Y más todavía si se piensa que la Agencia Estatal de Administración Tributaria (AEAT) española tiene una cantidad ingente de datos y es una de las más avanzadas del mundo en la utilización de técnicas de aprendizaje automático para la investigación tributaria. La base de datos Zujar contiene más de 20 billones

[14] Más recientemente, el Reino Unido ha puesto en marcha una estrategia nacional de datos (UK National Data Strategy, 2020).

de registros. Además, cuentan con Prometeo, Infonor, Midas y procedimientos avanzados de *web scraping*. Finalmente, los sistemas de inteligencia artificial de Electra o Teseo son capaces de generar resultados sorprendentes sobre relaciones comerciales de empresas y contribuyentes con imágenes de redes que permiten identificar de forma rápida si el patrón se corresponde con alguna de las tipologías habituales de fraude tributario.

Ciertamente las agencias tributarias han sido siempre muy refractarias a compartir sus datos incluso si se trata de realizar investigación con fines sociales. La AEAT no ha sido una excepción. Evidentemente la naturaleza de los datos recomienda ser muy cuidadosos en su gestión, aunque esto no debería impedir la realización de investigación imponiendo todo tipo de condiciones para garantizar que se cumplen estrictamente unos protocolos de anonimización y uso responsable de los datos. Pero en los últimos años la AEAT ha ido mostrándose más receptiva a compartir sus datos. La colaboración entre la AEAT y el INE en la construcción del Atlas de la Distribución de la Renta de los Hogares es un ejemplo de esta apertura. La Agencia Tributaria también participó muy activamente, y fue fundamental, en la construcción del Sistema Estatal de índices de Alquiler de Vivienda, que proporciona información de precios de alquiler hasta nivel de distrito censal. Otro signo de este nuevo tiempo es que la AEAT haya permitido a la AIReF utilizar las declaraciones de renta anonimizadas (35,5 millones) que se han cruzado con la Encuesta de Presupuestos Familiares para analizar la fiscalidad conjunta de IRPF e IVA de las familias (AIReF, 2020), aunque estos datos no son públicos.

En los últimos años también se han hecho avances en otros datos administrativos. La tradicional muestra continua de vidas laborales (MCVL) se ha enlazado con el IRPF. El PET (panel de datos de empresa-trabajadores) proporciona una visión similar a la MCVL pero desde la perspectiva de la empresa a partir del enlace de los registros sobre las empresas y las vidas laborales de los trabajadores. Asimismo, la creación del laboratorio de datos del Banco de España permite acceder a los datos de la Central de Balances. Dentro de las colaboraciones entre organismos públicos e instituciones de investigación el DataReSS, una colaboración entre el instituto de estadística de Cataluña (IDESCAT) y la Barcelona Graduate School of Economics, facilita el acceso a los registros que producen las administraciones públicas catalanas.

En todo caso, y con los avances, todavía queda mucho camino por recorrer. Como señala con acierto el documento de la AIReF (2020) «España no está aprovechando el potencial para la evaluación de las políticas públicas que se deriva de los datos administrativos en poder de las administraciones públicas». Experiencias recientes muestran que falta todavía mucho por hacer y que las propuestas de AIReF (2020) son un buen punto de partida para recorrer ese camino. De todas formas, en los últimos meses se ha producido dos hitos relevantes que potencialmente podrían ser muy transformadores. En primer lugar, se creó en agosto de 2020 la Oficina del Dato, dependiente de la Secretaria de Estado de Digitalización e Inteligencia Artificial. El artículo 2 del BOE de su creación señala cinco objetivos:

a) El diseño de las estrategias y marcos de referencia en materia de gestión de datos, la creación de espacios de compartición de datos entre empresas, ciudadanos y Administraciones Públicas de manera segura y con gobernanza (*sandboxes*, *data spaces* nacionales y europeos, ecosistemas de datos para uso sectorial tanto público como privado, etc.) y el empleo masivo de los datos en los sectores productivos de la economía mediante tecnologías *Big Data* e Inteligencia Artificial, entre otras, así como el desarrollo de mecanismos de acceso seguros a estas plataformas de datos, para la toma de decisiones públicas basadas en datos o para uso empresarial, garantizando su seguridad y gobernanza a través de arquitecturas API u otros mecanismos
b) El diseño de las políticas de Gobernanza y estándares en la gestión y análisis de datos que deben regir en la Administración General del Estado. Coordinación de modelos, recomendaciones y valoraciones sobre soluciones tecnológicas de codificación, anonimización y tratamiento de datos, sistemas de geolocalización, plataformas y modelos de intercambio, interacciones, modelizaciones y valoración de riesgos, seguridad en la gestión y almacenamiento de los datos, entre otros.
c) El desarrollo de un Centro de Competencia de analítica avanzada de datos que defina las metodologías y mejores prácticas y que asegure que se desarrollan las competencias tecnológicas y las herramientas necesarias para la toma de decisiones basadas en datos por parte de las Administraciones Públicas, permitiendo el desarrollo de políticas basadas en evidencia.
d) La formación y desarrollo de mecanismos de transferencia de conocimiento a los distintos ministerios y Administraciones Públicas.
e) La coordinación técnica de las iniciativas en materia de datos de los distintos departamentos ministeriales y Administraciones Públicas en el marco de las estrategias y programas de la Unión Europea.

El segundo hito relevante es el comunicado institucional del INE, la Agencia Tributaria, la Seguridad Social y el Banco de España (2021) para trabajar conjuntamente en el desarrollo de un sistema de acceso a sus bases de datos con fines científicos de interés público.

Conclusiones

La utilización de datos masivos y algoritmos de aprendizaje automático tienen cada vez un papel más relevante en la aproximación empresarial a las soluciones de mercado y la investigación con fines sociales. La colaboración entre autoridades públicas también puede favorecer la utilización y fusión de datos administrativos de diversas administraciones. Por su parte la creciente participación del sector privado en la generación de datos útiles para la investigación económica (evaluaciones de políticas públicas, seguimiento de la economía a alta frecuencia, etc.) hace cada

vez más importante la colaboración público-privada en el aprovechamiento de estas bases de datos. En este contexto, el acceso a microdatos bancarios proporciona una de las fuentes de información con mayor potencial. Así lo muestran multitud de estudios recientes que utilizan este tipo de datos para analizar, con gran granularidad y alta frecuencia, fenómenos económicos muy relevantes como las consecuencias de la pandemia de COVID-19 o el impacto de las políticas destinadas a atenuar sus efectos.

Hasta hace pocos años el aprovechamiento de los datos administrativos de las instituciones públicas españolas para la evaluación de políticas públicas ha estado muy alejado del enorme avance que se estaba produciendo en otros países. En los últimos tiempos se observa una mayor sensibilización de las instituciones públicas respecto a la importancia de los datos administrativos para la investigación económica, aunque todavía queda mucho camino por recorrer en la sensibilización del público y los medios de comunicación. Ejemplos de este cambio son la creciente colaboración de la Agencia Tributaria en la construcción de estadísticas derivadas de sus datos, la creación de la Oficina del Dato y el reciente anuncio de INE, la Agencia Tributaria, la Seguridad Social y el Banco de España de desarrollar un sistema de acceso a sus bases de datos con fines científicos de interés público. Aunque llevamos bastante retraso respecto a otros países estas iniciativas abren una ventana de oportunidad que, dependiendo de cómo se concrete, puede producir un avance muy significativo en la calidad y relevancia de la investigación económica en España.

Referencias bibliográficas

AIReF, Autoridad Independiente de Responsabilidad Fiscal (2020, octubre). *Opinión para una estrategia de acceso a datos administrativos*. Opinión 1/20.

Aspachs, O., Durante, R., Graziano, A., Mestres, J., Reynal-Querol, M., & Montalvo, J. G. (2021). Tracking the impact of COVID-19 on economic inequality at high frequency. *PloS one*, *16*(3). https://doi.org/10.1371/journal.pone.0249121

Athey, S., Ferguson, B. A., Gentzkow, M., & Schmidt, T. (2020). *Experienced segregation* (NBER Working paper No. 27572). National Bureau of Economic Research.

Baker, S. R. (2018). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, *126*(4), 1504-1557.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). *Income, liquidity, and the consumption response to the 2020 economic stimulus payments* (NBER Working paper No. 27097). National Bureau of Economic Research.

Barcellan, R., Nielsen, P., Calsamiglia, C., Camerer, C., Cantillon, E., Crépon, B., ... Wright, L. (2017). Developments in Data for Economic Research. In L. Matyas, R. Blundell, E. Cantillon, B. Chizzolini, M. Ivaldi, W. Leininger et al. (Eds.), *Economics without borders: Economic Research for European Policy Challenges* (pp. 568-611). Cambridge University Press. doi:10.1017/9781316636404.015

Bounie, D., Camara, Y., Fize, E., Galbraith, J., Landais, C., Lavest, C., & Savatier, B. (2020). Consumption dynamics in the covid crisis: real time insights from French transaction bank data. *Covid Economics*, *59*, 1-39.

Card, D., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *American Economic Association, ten years and beyond: Economists answer NSF's call for long-term research agendas*. National Science Foundation

Carvalho, V. M., Hansen, S., Ortiz, A., Garcia, J. R., Rodrigo, T., Rodriguez Mora, S., & Ruiz de Aguirre, P. (2020). *Tracking the COVID-19 crisis with high-resolution transaction data*. CEPR Discussion Paper No. DP14642.

Carvalho, P., Peralta, S. & dos Santos, J. P. (2020). *What and how did people buy during the Great Lockdown? Evidence from electronic payments*. ECARES Working Paper 2020-20.

Cavallo A., & Rigobon R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, *30*(2), 151-78.

Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, *60*(2), 152-165.

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*(21), 8589–8594.

Chen, S., Igan, D. O., Pierri, N., Presbitero, A. F., Soledad, M., & Peria, M. (2020). Tracking the economic impact of COVID-19 and mitigation policies in Europe and the United States. *IMF Working Papers*, *2020*(125).

Chetty, R., Friedman, J. N., Hendren, N., & Stepner, M. (2020). *Real-time economics: A new platform to track the impacts of COVID-19 on people, businesses, and communities using private sector data* (NBER Working Paper No 27431). National Bureau of Economic Research.

Chetty, R., Grusky D., Hell M., Hendren N., Manduca R., & Narang J. (2016). The Fading American Dream: Trends in Absolute Income Mobility Since 1940. *Science* 356(6336), 398-406.

Cox, N., Ganong, P., Noel, P., Vavra, J., Wong, A., Farrell, D., ... Deadman, E. (2020). Initial impacts of the pandemic on consumer behavior: Evidence from linked income, spending, and savings data. *Brookings Papers on Economic Activity*, *2020*(2), 35-82.

Davidowitz, S. (2017). *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really*. Dey Streets Books

Derman, E. (2004). *My life as a quant*. John Wiley & Sons.

Doerr, S., Gambacorta, L., & Garralda, J. M. S. (2021). Big data and machine learning in central banking. *BIS Working Papers*, (930).

Duhigg, C. (2011). *The power of habit*. Random House.

García Montalvo, J. (2003). Liquidity and market makers: a pseudo-experimental analysis with ultrahigh frequency data. *The European Journal of Finance*, *9*(4), 358-378.

García Montalvo, J. (2014). El impacto del big data en los servicios financieros. *Papeles de Economía Española*, *Número extraordinario*, 43-59.

García Montalvo, J., & Reynal-Querol M. (2020). *Distributional Effects of COVID-19 on Spending: A First Look at the Evidence from Spain*, Barcelona GSE Working Paper 1201.

García Montalvo, J. (2021). Data science y sus aplicaciones económicas: una perspectiva personal. En D. Peña, P. Poncela & E. Ruiz (Eds.), *Análisis Econométrico y Big Data* (pp. 5-24). FUNCAS.

García Montalvo, J., Reynal-Querol M., & Muñoz J.C. (2021). *Measuring inequality from above*. Barcelona GSE Working Paper No. 1252.

García Montalvo, J., & Reynal-Querol, M. (2021). Ethnic diversity and growth: Revisiting the evidence. *Review of Economics and Statistics*, *103*(3), 521-532.

Gelman, M., Kariv, S., Shapiro, M. D., Silverman, D., & Tadelis, S. (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science*, *345*(6193), 212-215.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012-1014.

Golec, P., Kapetanios, G., Neuteboom, N., Ritsema, F., & Ventouri, A. (2020). *Disentangling the effect of government restrictions and consumers' reaction function to the Covid-19 pandemic: evidence from geo-located transactions data for The Netherlands*. DAFM Working Paper Series, No. 2020/4.

Hacıoğlu-Hoke, S., Känzig, D. R., & Surico, P. (2021). The distributional impact of the pandemic. *European Economic Review*, *134*, 103680.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: a computational linguistics approach. *The Quarterly Journal of Economics*, *133*(2), 801-870.

Instituto Nacional de Estadística, Seguridad Social & Banco de España. (2021, 13 de abril). Comunicado oficial por el cual el Instituto Nacional de Estadística, la Seguridad Social y el Banco de España acuerdan trabajar conjuntamente en el desarrollo de un sistema de acceso a sus bases de datos con fines científicos de interés público. [comunicado de prensa].

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790-794.

Kubota, S., Onishi, K., & Toyama, Y. (2021). Consumption responses to COVID-19 payments: evidence from a natural experiment and bank account data. *Journal of Economic Behavior & Organization*, *188*, 1-17.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205.

Olafsson, A., & Pagel, M. (2018). The liquid hand-to-mouth: Evidence from personal finance management software. *The Review of Financial Studies*, *31*(11), 4398-4446.

Pinkovskiy, M., & Sala-i-Martin, X. (2016). Lights, camera… income! Illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, *131*(2), 579-631.

Sheridan, A., Andersen, A. L., Hansen, E. T., & Johannesen, N. (2020). Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. *Proceedings of the National Academy of Sciences*, *117*(34), 20468-20473.

Williamson, E., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., ... Goldacre, B. (2020). OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *MedRxiv*.

Yu, P., McLaughlin, J., & Levy, M. (2014). Big Data: A Big Disappointment for Scoring Consumer Credit Risk. *NCLC, National Consumer Law Center, Boston, MA*, *14*.

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

La evaluación como proceso de aprendizaje: una necesidad del presente, una inversión de futuro

Evaluation as a learning process: a present-day necessity, a future investment

Sara de la Rica Goiricelaya
Lucía Gorjón García
Imanol Lizarraga Álvarez
Fundación ISEAK

Resumen

En la era del Big Data, una verdadera cultura de la evaluación de las políticas públicas se muestra al alcance de la mano. En el presente documento se plantea el papel que juega la evaluación de políticas públicas en la mejora de las administraciones y la resolución de algunos de los retos socioeconómicos más acuciantes para nuestras sociedades. Se incide asimismo en el gran salto adelante que supondría una mejora del uso de las bases de datos administrativas para la evaluación. De modo ilustrativo, se presenta un estudio de caso para una mejor comprensión de los beneficios potenciales y tangibles de estas inversiones en apertura y conocimiento.

Palabras clave: *evaluación, políticas públicas, retos sociales.*
Clasificación JEL: *I31, I38, D31, D63.*

Abstract

With the rise of Big Data, a genuine culture of evaluation has become within reach. This article discusses the importance of public policy evaluation in the improvement of public administrations and the resolution of some of our societies' most urgent socioeconomic concerns. Additionally, it underlines the enormous leap forward that an increase in the utilization of administrative data for evaluation would imply. Finally, to illustrate the potential and tangible benefits of those investments in openness and knowledge, a case study is presented.

Keywords: *evaluation, public policy, social challenges.*

DOI: https://doi.org/10.32796/cice.2021.102.7312
Recibido: Agosto 2021 · Aceptado: Septiembre 2021

1. Introducción

El debate sobre la evaluación de políticas públicas se ha rearmado en los últimos años, situándose entre las principales prioridades de numerosos gobiernos e instituciones a todos los niveles. Detrás de este creciente interés se encuentran la mejora de la computación, de las bases de datos y de las técnicas apropiadas para una evaluación precisa, así como una mayor concienciación acerca del buen uso de los fondos públicos en aras de una mayor transparencia y rendición de cuentas hacia la ciudadanía. Las posibilidades para perseguir una estrategia de evaluación ambiciosa e innovadora son así mayores que nunca.

Todo programa o política pública se pone en marcha con la intención de resolver problemas y paliar necesidades de la sociedad a la que concierne. No obstante, sin un seguimiento preciso de sus objetivos y resultados, cualquier intervención corre el riesgo de fracasar en su cometido. Sin soluciones adecuadas, los problemas se enquistan y complican. Más allá de buenas intenciones, urgen buenas actuaciones.

La evaluación de políticas públicas se ha popularizado en los últimos años como una forma de abordar esta necesidad. Nos permite guiar el buen funcionamiento de las políticas, destacando las buenas prácticas y advirtiendo de los puntos a mejorar. De este modo, aporta conocimiento útil para la toma de decisiones, mejorando la eficacia y la eficiencia de los programas. En esta época de grandes retos sociales y con una ciudadanía cada vez más concienciada, la exigencia ante las intervenciones públicas es notable.

Para una definición más formal del concepto de evaluación de políticas públicas, nos remitimos aquí a la otorgada por el Ministerio de Política Territorial:

> La evaluación de políticas públicas es un proceso sistemático y razonado de generación de conocimiento, a partir de la recopilación, análisis e interpretación de información, encaminado a la compresión global de una intervención pública –sea esta una política, plan, programa o norma–, para alcanzar un juicio valorativo, basado en evidencias, respecto de su diseño, puesta en práctica y efectos (resultados e impactos). (IEPP, 2021)

Como se indica, la evaluación puede enfocarse en cualquiera de las numerosas fases de determinado programa o estudiar el mismo de forma íntegra. Así, una pertinente evaluación ayuda a reconocer si los objetivos planteados están bien definidos, si el diseño del programa se ajusta a estos y si los resultados son satisfactorios.

En este sentido, la evaluación debe verse como una inversión: permite racionalizar el gasto de las diversas administraciones públicas, mejorando su eficacia, eficiencia y flexibilidad; respecto a la ciudadanía a la que sirve, dota a los cuerpos públicos de una mayor transparencia y rendición de cuentas. Mejores actuaciones acercan las instituciones públicas a la sociedad. Adicionalmente, en un entorno cambiante y complejo, las buenas decisiones han de asentarse en la evidencia, subrayando así el papel de la evaluación como aliada natural de la acción pública. En definitiva, la

cultura de la evaluación es un paso más en la dirección hacia una sociedad de compromiso, abierta al aprendizaje y a la evidencia.

Por otra parte, ninguna evaluación es viable sin buenos datos. La información rica y completa es el mejor recurso de unas administraciones ambiciosas en sus objetivos. A la respuesta de cómo evaluar cabe dar, por tanto, una respuesta corta y directa: reuniendo datos suficientes, primero, y utilizando las técnicas adecuadas, en segundo lugar. Partiendo de estas dos premisas, prácticamente cualquier programa o política es potencialmente evaluable, bien de forma externa, bien internamente.

En el presente estudio se remarca el papel de la evaluación como un aprendizaje continuo, dinámico y con gran potencial para mejorar la acción pública. En primer lugar, se explicará brevemente el concepto de evaluación, dando ligeras pinceladas sobre su importancia y potencial. A continuación, se debatirá sobre el posible alcance de la evaluación –qué se puede evaluar y cómo hacerlo–. Se mostrará el modelo de evaluación integral, que permite estudiar cualquier política o programa desde su fase de diseño hasta los resultados, pasando por la implementación y el seguimiento. En tercer lugar, se introducirá en el debate la importancia del uso de bases de datos administrativos para la evaluación, haciendo especial hincapié en registros administrativos, en particular en las bases de la Seguridad Social. Para ilustrar la aplicación de lo anteriormente expuesto, se presenta en el cuarto apartado un estudio de caso reciente donde se evaluó el contrato en prácticas profesionales en España. Se cierra el estudio con unas breves conclusiones y recomendaciones sobre el camino a tomar.

2. Qué y cómo evaluar: la evaluación integral

Con datos suficientes, cualquier intervención pública es evaluable, desde el impacto de un programa de empleo para determinado colectivo hasta una estrategia de lucha contra la pobreza. Para ello, es preciso abordar una inversión inicial en aprendizaje, que permitirá ahorrar cantidades de tiempo, dinero y esfuerzos potencialmente importantes en un futuro. No obstante, el principal y mayor objetivo debería ser siempre aprender sobre las intervenciones realizadas, sobre su alcance, sus limitaciones y sus posibles mejoras. La evaluación puede extenderse a cualquier ámbito de actuación (también privado) que sea medible, como estrategias de gestión o mejoras organizativas. No obstante, es en las políticas públicas, que abordan necesidades sociales crecientes con recursos limitados, donde la evaluación tiene una repercusión de mayor calado (AEVAL, 2010).

Si bien los métodos de evaluación son numerosos, la evaluación integral se considera la más completa y profunda. Se trata de un procedimiento que persigue una comprensión más amplia de las distintas fases en que se desarrolla un programa o política y una identificación exacta de los problemas que puedan llevar a una desviación de los resultados buscados. En este sentido, en lugar de centrarse exclusivamente en la pertinencia de la política en cuestión o en sus resultados, estudia todas y cada

una de las fases, desde el diseño hasta el impacto, pasando por la implementación y el seguimiento.

Las evaluaciones integrales pueden comenzar antes de la implementación del programa *(ex ante),* mientras este se ejecuta (intermedia) o al concluir *(ex post)*. Cada enfoque pone el acento en una de las fases de la evaluación, si bien la única diferencia radica en el momento en que se aborda la misma. Mientras que la evaluación intermedia o *ex post* solo permiten la corrección de errores para futuras implementaciones del mismo programa o similares (por realizarse con el programa ya en marcha), la evaluación *ex ante* posibilita la detección temprana de puntos débiles y la resolución dinámica de problemas sobre la marcha. En efecto, cuanto antes se comience la misma, antes se podrán corregir los posibles fallos de diseño que pudieran poner en peligro los resultados demandados. Si bien tradicionalmente las evaluaciones intermedias y *ex post* han estado más extendidas, una buena cultura de la evaluación debería dar mayor valor a las evaluaciones *ex ante,* que permiten un mejor diseño de políticas públicas y una corrección pronta de problemas de base (AEVAL, 2010).

Como se explicará en las secciones siguientes, la evaluación integral persigue responder diferentes preguntas en cada estadio de la política. En la primera etapa de la evaluación, las cuestiones más relevantes radicarán en la pertinencia del programa y la credibilidad de los resultados esperados. En la segunda etapa, se pondrá el acento en si la implementación está poniéndose en práctica de la forma planeada; en la tercera, se estudiará si los resultados medibles están a la altura de los objetivos planteados o si bien la política tuvo unas consecuencias inesperadas y posiblemente indeseadas. Finalmente se analizará el impacto de la política en cuestión.

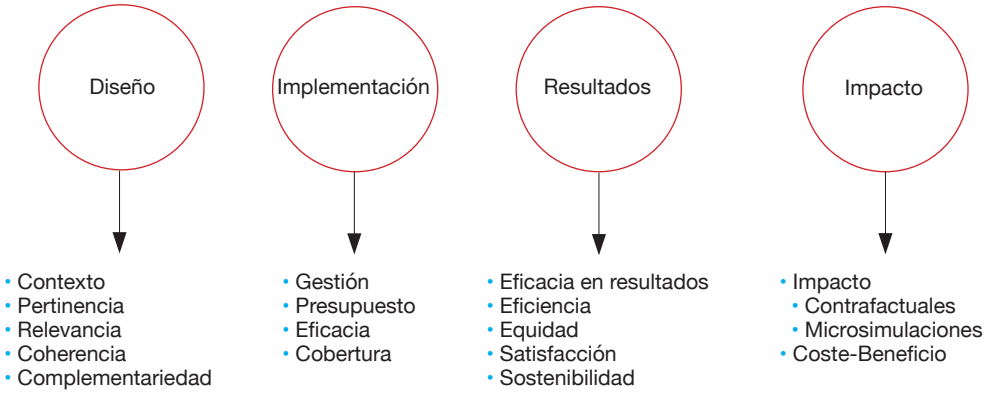
2.1. Las cuatro etapas de la evaluación integral

Como se introdujo previamente, la evaluación integral sigue una metodología que consta de cuatro etapas, donde se estudian cada una de las fases del ciclo de una política o programa: diseño, implementación, resultados e impacto. Para cada una de las fases se plantea una serie de preguntas y subpreguntas, hipótesis que se tratan de confirmar.

Primera etapa: el diseño es el primer paso en la elaboración de políticas o programas. Se eligen aquí los objetivos a cumplir, se establece la cadena lógica que llevará a su consecución y se dota al conjunto de un marco estratégico con plazos fijos y metas cuantificables. La evaluación de esta fase inicial permite detectar de forma anticipada estrategias desenfocadas y objetivos confusos. Si se evalúa el diseño de un programa aún no implementado, una intervención temprana favorece su refinamiento y reorientación (Instituto para la Evaluación de Políticas Públicas, 2020a).

Son varios los criterios a aplicar a propósito del diseño. En primer lugar, debe haber un diagnóstico del contexto socioeconómico donde se aplicará la política en

FIGURA 1
ETAPAS Y CRITERIOS DE LA EVALUACIÓN INTEGRAL



FUENTE: Elaboración propia.

cuestión. Sin conocer el entorno socioeconómico, así como las fortalezas y debilidades existentes, la política corre el riesgo de estar alejada de los problemas y posibilidades reales. Para este ejercicio de contextualización se puede recurrir tanto a bases de datos públicamente disponibles a través de instituciones como los Institutos de Estadística nacionales o regionales, como a registros administrativos, que aportan una información rica en detalle sobre la población en la que intervenir. Segundo, una vez fijado el entorno, cabe realizar un análisis de pertinencia. Las condiciones cambiantes pueden provocar que los objetivos inicialmente fijados queden alejados de los problemas reales. En tercer lugar, ha de estudiarse la relevancia de la política en cuestión. Para ello, se analizará su importancia en la resolución de problemas y se comparará con otras políticas de similares intenciones. El objetivo es la búsqueda de las medidas con capacidad tractora que mayor impacto tengan, alcanzando así en mayor medida los objetivos. Cuarto, se debe hacer hincapié en la coherencia del programa, estudiando los recursos –tanto presupuestarios como humanos–, las medidas que se implementarán y la cadena lógica que generará los resultados. Un programa con objetivos más específicos, financiación suficiente y una estrategia coherente es preferible a uno con ideas más amplias y vagas sobre cómo se conseguirán los objetivos deseados. En quinto y último lugar, se debe atender a la posible complementariedad con programas que afecten a las mismas necesidades y objetivos, con intención de esquivar duplicidades.

Más allá de los anteriores análisis, el uso de microsimulaciones puede ser de especial interés para comprender desde la etapa inicial si el diseño de la política llevará a los resultados esperados. Para ello, pueden realizarse pruebas piloto, donde algunas personas reciban la intervención pública de forma experimental, sirviendo su comportamiento y los resultados observados para determinar la viabilidad de la política,

estimar el coste de su puesta en marcha y simular los efectos de la acción a mayor escala, siempre bajo determinados supuestos que deben estar razonados en estudios académicos previos o un análisis riguroso del entorno en que operar. Las microsimulaciones no solo permiten estimar el coste de la política, sino también identificar el número óptimo de usuarios potenciales o priorizar aquellas acciones con efectos positivos más significativos.

Segunda etapa: tras el periodo de diseño, la siguiente etapa en el ciclo es la implementación. La evaluación en esta fase se centra en la operativa y organización de la puesta en marcha y seguimiento del programa. Es de especial interés una verificación exhaustiva de la ejecución de los fondos presupuestarios destinados al programa, así como de la gestión y funcionamiento interno del mismo. Las dificultades para gastar los fondos en el lugar adecuado o la escasez de personal pueden hacer descarrilar programas perfectamente válidos. A este respecto, deben estudiarse las posibles barreras de cualquier índole que entorpezcan el programa. La selección de personal del perfil apropiado para las tareas a realizar es asimismo crucial. La calidad de la información que se esté recogiendo es, no obstante, la preocupación primordial en estas etapas iniciales. Solo mediante bases de datos suficientes será posible la evaluación posterior de resultados e impacto. Tras la evaluación del funcionamiento del programa, se debe atender a la eficacia en cuanto a la correcta consecución de fases, procedimientos y objetivos intermedios (con independencia del coste de ejecución). Para ello, se evalúan el grado de cumplimiento y la consumación en los plazos establecidos. Unido al anterior criterio, por último, se encuentra el análisis de cobertura (Instituto para la Evaluación de Políticas Públicas, 2020b). Mediante el mismo se pretende contabilizar tanto la proporción de la población objetivo que ha sido alcanzada por la política –y compararla con los objetivos iniciales–, como la proporción de personas fuera del grupo objetivo que participa en el programa –una medida de lo bien enfocada que está una política y del modo ineficiente en que se gastan los fondos públicos.

Tercera etapa: con el programa o política a pleno rendimiento, cabe remitirse a la evaluación de resultados, la tercera etapa de la evaluación integral, donde se intenta responder hasta qué punto han sido logrados los objetivos de la política o programa. En esta etapa, no se busca cuantificar específicamente el posible impacto de la política, sino analizar si se han generado los resultados esperados. Por definición, se trata de una evaluación *ex post*. Son cinco los criterios a seguir en la evaluación de resultados: la eficacia en resultados, la eficiencia, la equidad, la satisfacción y la sostenibilidad. La eficacia en resultados hace referencia al grado de cumplimiento del programa. La eficiencia, mientras, se refiere a la cantidad de recursos empleados para conseguir (o perseguir) los objetivos iniciales. Equidad y satisfacción, por otro lado, se remiten a la igualdad de trato entre los diversos grupos y su conformidad con los resultados. Finalmente, la sostenibilidad del programa debe garantizar su autonomía para mantener los resultados en el tiempo (Instituto para la Evaluación de Políticas Públicas, 2020c).

Cuarta etapa: el último paso en la evaluación integral de una política es la evaluación de impacto. En este ciclo, se trata de atribuir resultados específicos a las acciones emprendidas dentro de la política, cuantificarlos y calibrar el éxito o fracaso en la consecución de objetivos. En definitiva, en la evaluación de impacto se busca una relación causal entre el programa planteado y los resultados obtenidos. Esta última fase de evaluación se ilustrará mediante un caso práctico en el apartado cuarto del presente estudio. Sin una base de datos adecuada, la evaluación de impacto resulta del todo imposible. Por ejemplo, en la evaluación de un programa de inserción para jóvenes, se debe disponer de datos sobre características sociodemográficas de beneficiarios y no beneficiarios, características sobre contratos y condiciones laborales previas, intermedias y posteriores, etcétera. Estos datos podrían permitir computar el impacto específico de la política en su inserción laboral, así como observar qué grupos socioeconómicos se ven más beneficiados y dónde se podría mejorar (AEVAL, 2010).

La evaluación de impacto puede realizarse bien mediante la construcción de contrafactuales, bien a través de microsimulaciones. Óptimamente, el impacto podría computarse comparando los resultados bajo la aplicación de la política y los resultados sin la misma. No obstante, en la vida real solo se observa la primera. Por ello, para ver el impacto de la política en el grupo de tratamiento (los participantes) se requiere de un grupo de control con quien poder comparar. El grupo de control debe tener unas características poblacionales lo más similares posibles al grupo de tratamiento, de tal modo que su única diferencia radique en la participación o no participación en el programa a evaluar. Comparando ambos y controlando por sus características puede estimarse de forma precisa y rigurosa el impacto real de la política. La evaluación de impacto puede ser experimental o cuasi-experimental según el método de selección de grupos de tratamiento y control. Respecto a la primera, se refiere a la realización de un experimento controlado, donde se selecciona aleatoriamente a participantes y no participantes. Cuando la población objetivo es pequeña, la evaluación experimental puede realizarse tomando a todos los potenciales individuos como muestra; cuando la población es grande, se toma una muestra representativa del grupo general. Las evaluaciones cuasi-experimentales**,** en cambio, se emplean cuando los experimentos son imposibles o moralmente reprobables –por ejemplo, cuando se necesita excluir a personas de la participación en un programa de vuelta al empleo para poder compararlos con personas que sí acceden–. En esta ilustración, los propios desempleados serían libres de elegir su participación o no en el programa, pudiendo llevar a problemas de sesgo de selección (donde las características de grupo de tratamiento y control fueran diferentes). Existen diversos métodos cuantitativos comúnmente extendidos, como el control de diferencias en diferencias, el emparejamiento o la regresión discontinua, que requieren cierto conocimiento técnico, pero aportan resultados robustos muy relevantes.

Siguiendo el proceso de evaluación en cada una de sus etapas se deben identificar las buenas prácticas, descartando o modificando aquellas políticas que por su

diseño o sus pobres resultados devienen inapropiadas y potenciando las de resultados positivos. La evaluación, no obstante, no es un proceso estático, sino una técnica dinámica en que el aprendizaje y el cambio razonado mejoren continuamente las intervenciones públicas.

3. La riqueza y potencial de las bases de datos administrativas

El combustible primordial a lo largo de todas las fases de la evaluación son los buenos datos. Sin medidas cuantificables, cualquier evaluación es inviable. Las fuentes a través de las cuales se pueden obtener los datos necesarios son numerosas: según su origen (de instituciones públicas o privadas), según su método de obtención (datos de registro, de encuestas, de experimentos, cuestionarios…) o según su duración en el tiempo (datos de sección cruzada, longitudinales o series temporales…).

No obstante, la gran riqueza de las diversas administraciones públicas en cuanto a los datos administrativos se refiere las hace protagonistas en la construcción de una cultura de evaluación moderna. Las administraciones recopilan una gran cantidad de datos de forma precisa que podrían ser de excepcional uso para la evaluación de programas públicos. Adicionalmente, las bases de datos administrativas son complementarias a las encuestas o experimentos. Tienen la gran ventaja de que abarcan a toda la población, por lo que los posibles problemas de tamaño que suelen acompañar a muchas bases de datos provenientes de encuestas no existen. El inconveniente es que a menudo no se dispone de toda la información que sería deseable para abordar una evaluación robusta.

Sin embargo, habitualmente no se reconoce su potencial. Más aún, la falta de integración de las diferentes bases de datos –incluso dentro de una misma institución–, dificulta su uso y lastra su utilidad. Sería recomendable una estrategia global que permitiera una mayor armonización entre las distintas fuentes de datos, con códigos y formas de medición comparables e integrables que salvaguarden el anonimato de las personas que aparecen en la base de datos. Dado que cada base de datos surge con un propósito diferente, una fácil integración entre las mismas permitiría unir perspectivas hasta obtener una comprensión más global.

Un buen ejemplo de buenas prácticas sobre el que avanzar sería la Base de Datos Integrada para la Investigación del Mercado Laboral (IDA), realizada por el servicio estadístico nacional de Dinamarca en colaboración con numerosas instituciones y administraciones del país, y que se estudia en mayor profundidad en Brugarolas *et al*. (2020). El servicio estadístico danés cuenta en total con más de 250 fuentes estadísticas distintas a disposición del público investigador que se pueden combinar según las necesidades del solicitante. Desde datos poblacionales hasta registros de pago de impuestos, cotizaciones a la seguridad social o datos sobre empleados de empresas, las posibilidades son numerosas. Los datos se ofrecen anonimizados y bajo un compromiso de confidencialidad que garantiza la buena fe de ambas partes, con el espíritu del empleo de las numerosas bases de registro público para la

investigación y la promoción de la evaluación de políticas. Es en este contexto que el servicio de estadística danés ofrece la IDA como una herramienta ágil y de gran utilidad para la investigación. Se trata de una base de datos anual donde se integran informaciones de diversas fuentes administrativas, obteniendo así en una misma base de datos información de los individuos, de los trabajadores y de las empresas. Respecto a los primeros datos, se incluyen abundantes estadísticas sociodemográficas, de residencia, condiciones familiares, sobre su experiencia profesional o los años en desempleo; en cuanto a los trabajadores, se proporciona información sobre sus periodos de empleo, sobre las condiciones laborales, tipo de contrato, jornada o salario hora; finalmente, se ofrecen datos sobre la localización de empresas, el número de trabajadores o la actividad. Huelga destacar la riqueza de esta base de datos, integrada por información ya existente de forma previa, pero que por su dispersión y falta de integración no tenía el potencial que ahora presenta. Adicionalmente, la IDA permite una gran flexibilidad para incorporar bases de datos externas de terceras fuentes, por lo que sus posibilidades son abiertas y numerosas. Kleven *et al*. (2019), por ejemplo, utilizan la IDA junto con datos de ingresos familiares del Ministerio de Finanzas danés para estudiar el impacto de la maternidad en la participación laboral de las mujeres. En definitiva, una pequeña inversión en recogida e integración de bases de datos dispersas ha permitido una gran expansión de la cultura de evaluación en el país.

Si se quisiera replicar una propuesta similar en España, no habría que empezar de cero. Existen diversas bases de datos muy ricas en información que podrían servir como columna vertebral de un proyecto de integración de bases administrativas ambicioso y de largo alcance. Un ejemplo de interés de base de datos administrativa de uso habitual para la evaluación de políticas –principalmente en el ámbito laboral– es la Muestra Continua de Vidas Laborales (MCVL). Esta base de datos consiste en la vida laboral completa de un 4 % de las personas afiliadas a la Seguridad Social. En particular, la base de datos recoge todos los episodios de empleo y desempleo de asalariados, con las fechas exactas y su duración, así como el tipo de contrato, horas trabajadas o tamaño de la empresa. También recoge información sobre el autoempleo, fecha de creación de la empresa, sector de actividad, ocupación o características sociodemográficas, así como información disponible sobre la base y grupo de cotización de cada una de las personas y las empresas. Su potencial es grande, pero la departamentización de las diversas bases de datos según la administración responsable dificulta notablemente su integración con otros datos sobre empresas o la población activa. Dado que los datos administrativos ya están recopilados para la gestión de la Seguridad Social, poner a disposición de la comunidad científica un método para su integración con otras bases de datos permitiría que se realizaran innumerables análisis de diversa índole en lo referente a la implementación de políticas laborales, de pensiones, de género, etcétera.

En este punto, es necesario destacar el reciente impulso a una estrategia de integración de bases de datos de diversas fuentes administrativas que está siendo liderada por el Instituto Nacional de Estadística (INE), mediante la cual se pretende

combinar datos del INE con información fiscal del Ministerio de Hacienda, así como de la Seguridad Social. Este interés y esfuerzo muestra que se avanza en el buen sentido, si bien el trabajo a hacer es aún ingente.

En el estudio de caso presentado a continuación, se muestra el impacto de un buen uso de bases de datos para la medición del impacto de una política activa de empleo dirigida, en concreto, al colectivo de personas jóvenes: los contratos en prácticas. Por otro lado, también se explicarán sus limitaciones, así como las posibilidades que podría abrir una propuesta de integración de bases de datos como la IDA de Dinamarca.

4. Estudio de caso: el contrato en prácticas. Cómo evaluamos los beneficios/perjuicios del contrato en prácticas

En anteriores apartados se incidió en la necesidad de entender la cultura de la evaluación como inversión de futuro. Para ilustrar el proceso y sus posibilidades se presenta a continuación un caso práctico de evaluación de una política pública de España, el contrato en prácticas profesionales para personas jóvenes (De la Rica *et al*., 2020).

En este caso en concreto, nos centramos en la última etapa de la evaluación integral, el estudio de impacto. No obstante, se deberían analizar asimismo tanto el diseño como la implementación y los resultados del contrato en prácticas para obtener una panorámica más amplia. Adelantando las conclusiones que se mostrarán posteriormente, cabe atribuir a un mal diseño e implementación del referido contrato los pobres resultados observados en la evaluación de impacto. Una evaluación precisa en las primeras etapas de la política en cuestión podría así haber evitado algunos de los problemas que se detectan en el estudio de impacto.

Originalmente ideado para impulsar una transición laboral adecuada de las y los jóvenes desde su etapa formativa superior hacia su etapa laboral, mediante el contrato en prácticas la empresa recibe ciertos beneficios fiscales a cambio del contrato de personas jóvenes, de modo que estas adquieran una práctica adecuada a su formación. La práctica profesional debe venir acompañada de una inversión relevante en formación por parte de la empresa, de tal modo que se generen los incentivos adecuados para la transición a la contratación indefinida tras la conclusión de las prácticas. A cambio, las empresas se ven beneficiadas mediante una reducción del 50 % de las cotizaciones a la Seguridad Social por el trabajador (incentivo retirado en 2018), así como de un salario menor al estipulado por el convenio para el puesto en cuestión durante el periodo subvencionable. El contrato en prácticas debe tener una duración mínima de 6 meses y un máximo de 2 años, con jornada completa o parcial, y la retribución no puede ser inferior al 60 % (durante el primer año) o un 75 % (el segundo año) del salario establecido por convenio para un puesto con esas mismas responsabilidades y tareas.

En definitiva, se persigue una incorporación fluida y rápida de las personas jóvenes al mercado laboral tras el fin de sus estudios, así como una mayor formación por parte de las empresas, generando así los incentivos para fomentar la estabilidad laboral y la formación adecuada de las personas jóvenes trabajadoras.

Para comprobar si los objetivos descritos para este tipo de contrato se cumplen, se evalúan de forma cuantitativa dos dimensiones particulares de este contrato en prácticas: en primer lugar, la permanencia en la empresa tras la extinción del mencionado contrato; segundo, la transición del contrato en prácticas a contrato indefinido, bien sea en la misma empresa, bien en otra. Para ello, se utiliza la referida Muestra Continua de Vidas Laborales (2017). Se trata de una base de datos muy rica en información relativa al historial laboral de una muestra de personas trabajadoras de España, con detalles sobre cada episodio de empleo y desempleo de los individuos incluidos en ella desde el comienzo de su vida laboral hasta el 31 de diciembre de 2017. Para cada episodio de empleo, se especifica el tipo de contrato y la duración del mismo, así como las características sociodemográficas, el nivel formativo, el género o el lugar de residencia.

Para realizar la evaluación, se selecciona de la MCVL a las personas menores de 30 años con estudios de Formación Profesional o universitarios (requisito para firmar el contrato en prácticas) y se los divide en dos grupos, que servirán para hacer las comparaciones y estimar resultados cuantitativos –números interpretables, en definitiva–. El primer grupo, el grupo de tratamiento, lo componen personas jóvenes cuya primera relación laboral fue un contrato de prácticas; el segundo, el grupo de control, lo constituyen jóvenes de mismas características que los anteriores, pero cuya primera relación laboral fue un contrato temporal de más de tres meses (el primer contrato más común entre las personas jóvenes que entran al mercado laboral a través de un contrato temporal). Como se observa, el grupo de tratamiento es aquel que participa en la política a evaluar, mientras el grupo de control lo forman las personas con quienes compararemos las trayectorias laborales de los primeros. Si los grupos son similares en cuanto a sus características sociodemográficas, pero se observan divergencias en sus trayectorias posteriores, puede inferirse que las diferencias son imputables al impacto de la política en cuestión, el contrato en prácticas. En este sentido, estamos en posición de evaluar los dos objetivos planteados inicialmente: la permanencia en la empresa y la transición a indefinido.

Respecto al primer objetivo, dado que las empresas realizarían una inversión en formación del trabajador, se esperaría que la permanencia tras las prácticas fuera superior (o al menos no inferior) a las personas jóvenes con contrato temporal ordinario (grupo de control). La teoría económica indica que, cuando las empresas invierten en el capital humano de sus trabajadores/as, esperan retenerlos el tiempo suficiente para recuperar su inversión inicial a través de una mayor productividad del trabajador. En contra del espíritu original de la política, no obstante, se encuentra que quienes firman un contrato en prácticas tienen mayor probabilidad de irse de la empresa que aquellas personas jóvenes cuya primera experiencia laboral es un contrato temporal estándar de más de tres meses. Entre las personas jóvenes que

completan todo el periodo estipulado de prácticas, solo un tercio se queda con un contrato distinto tras la extinción de las prácticas, una proporción similar a la de las personas del grupo de control, que trabajan bajo contrato temporal. En cambio, entre aquellas que rescinden el contrato en prácticas con anterioridad, tan solo un 16 % se queda en la empresa, frente a un 73 % que cambia de empresa y empleo y un 11 % que transita hacia el desempleo. Teniendo en cuenta que más de la mitad de los contratos de prácticas se extinguen fuera del periodo estipulado, la tasa de permanencia resulta notablemente más baja entre las personas jóvenes en prácticas, respecto a aquellas personas con contrato temporal. Es decir, el contrato en prácticas no estaría fomentando la permanencia en la empresa, sino el efecto contrario. Este resultado sería muy consistente con el hecho de que las empresas no estarían invirtiendo en la formación de estos jóvenes y, por tanto, no existiría el incentivo de mantenerlos en su empresa tras las prácticas subvencionadas.

Segundo, respecto a la transición hacia un contrato indefinido, se observa que aquellas personas que permanecen en la empresa tras finalizar el contrato en prácticas tienen una probabilidad un 20 % inferior de conseguir un contrato indefinido tras las prácticas que quienes empiezan con contrato temporal. Por otro lado, entre quienes finalizan el contrato de prácticas fuera del periodo estipulado y se cambian de empresa, la probabilidad de obtener uno indefinido se incrementa en un 15 %. De nuevo, si las empresas formaran al/la trabajador/a, tendrían el incentivo de retener a esas personas mediante la oferta de un contrato estable al finalizar las prácticas. En ausencia de formación, sin embargo, la barrera que suele suponer a una empresa el salto desde un contrato de prácticas a uno indefinido se haría efectiva y las empresas preferirían ofrecer la alternativa que pase por la renovación bajo un contrato temporal estándar (algo que está permitido por ley durante otros dos años) o la extinción de la relación contractual. Adicionalmente, se encuentra un efecto indirecto y no esperado del contrato en prácticas y es que la probabilidad de obtener un contrato indefinido es superior entre aquellas personas jóvenes que no acaban el periodo estipulado de prácticas y se marchan de la empresa en cualquier punto intermedio. Así, una interpretación posible es que las personas jóvenes contratadas mediante un contrato de prácticas, conscientes de la falta de formación y la baja probabilidad de transición hacia un contrato indefinido en la empresa utilizarían el contrato en prácticas de manera temporal a la espera de mejores ofertas laborales. Otra posible explicación podría recaer en la exigencia de experiencia previa para muchas ofertas laborales. Así, las personas jóvenes podrían obtener la experiencia requerida a través del periodo de prácticas, con intención de incrementar sus oportunidades en futuros procesos de selección en terceras empresas.

Los anteriores resultados permiten comprender el impacto que el contrato en prácticas ha tenido en los últimos años sobre las carreras profesionales de las personas jóvenes. En una suerte de círculo vicioso, los jóvenes lo utilizan como contrato puente mientras buscan oportunidades laborales en otras empresas, ya que no tienen evidencia de que la empresa quiera retenerles; las empresas, mientras, saben que los jóvenes lo utilizan como empleo provisional, no invierten en su formación y lo

utilizan como una forma de obtener beneficios fiscales que abaratan el coste laboral. En resumen, el uso que se hace del contrato en prácticas es contrario al espíritu con el que fue creado, incrementando aún más la precariedad laboral de las personas jóvenes.

4.1. Qué aprendemos

La evaluación se puede definir sucintamente como un aprendizaje continuo sobre los factores que no se han tenido en cuenta en el diseño de políticas de cualquier índole. Como se ha demostrado en el apartado anterior, el diseño e implementación del contrato en prácticas no previó determinados comportamientos de empresas y trabajadores/as que han llevado a resultados diferentes a los esperados. Las buenas intenciones son necesarias, pero los buenos resultados lo son más.

La evaluación saca a la luz determinados aspectos a cambiar para conseguir un contrato en prácticas/formativo que sea eficaz en su propósito fundamental. Cualquier contrato en este sentido debe llevar aparejados ciertos requerimientos acreditables de formación en el puesto de trabajo que aseguren la implicación formativa de la empresa en la trayectoria profesional del/la trabajador/a. Por otro lado, las personas jóvenes deben ver beneficios tangibles en las prácticas, de tal modo que su compromiso con la empresa lleve a una maximización de la utilidad que la formación pueda aportar a sus aptitudes profesionales. Construir sobre el aprendizaje que aportan evaluaciones como la presentada previamente debería ser una prioridad para cualquier administración o institución que aspire a servir adecuadamente a su papel social.

4.2. Qué falta

Por supuesto, el estudio del contrato en prácticas referido deja numerosas zonas oscuras que los datos no permiten discernir. Una mejora de la gestión de bases administrativas podría permitir suplir esas faltas y alcanzar resultados más precisos, que ayudarían a orientar las políticas públicas hacia una más eficiente consecución de sus objetivos.

Una base de datos que integrara la MCVL con datos sobre empresas, por ejemplo, permitiría conocer de primera mano las diferencias entre las empresas que mejor están empleando el contrato en prácticas y aquellas que no lo usan adecuadamente. Adicionalmente, esta permitiría conocer el perfil de la empresa y el tipo de contrato que ofrece a otros empleados, la importancia de la cultura de la formación o la estabilidad laboral de sus trabajadores/as. Por otro lado, una mayor integración de la MCVL con otras fuentes de datos sociodemográficas o de Hacienda permitiría conocer la evolución de los ingresos de los jóvenes que firmaron un contrato en prácticas, su campo de estudio o su desempeño académico, todo ello vital para una mejor comprensión del mundo laboral que afrontan los jóvenes de hoy.

En resumen, los datos están ahí, solo se necesita un impulso público atrevido que garantice el ecosistema adecuado para una cultura de la evaluación rica, moderna e inteligente.

5. Hacia dónde ir

En un mundo informatizado con una creciente cantidad de datos, donde los grandes retos socioeconómicos y demográficos requieren soluciones cada vez más complejas, las administraciones e instituciones públicas debieran dirigirse hacia una mayor integración y mejor gestión de la gran riqueza que poseen: los datos. Como se mostró con el modelo danés, pequeñas inversiones pueden deparar a medio y largo plazo grandes rendimientos y mejoras de eficiencia notables. Las instituciones españolas cuentan con grandes profesionales que pudieran poner al país a la vanguardia de la transparencia y la rendición de cuentas pública.

Adicionalmente, la gran competencia por los fondos públicos en el contexto de un Estado de Bienestar en proceso de adaptación a la economía globalizada actual requiere una mejora continua de la eficacia y eficiencia con que se asignan los recursos públicos y se diseñan las políticas públicas que permitan un mayor bienestar social. La evaluación integral de políticas es nuestra mejor herramienta para lidiar con problemas que aparentemente carecen de solución, como la debilidad de la infraestructura laboral, los problemas en el sistema educativo o, más recientemente, los puntos flacos detectados en la sanidad y gestión pública a raíz de la pandemia de Covid-19. Además, una cultura de la evaluación que valore la ecuanimidad, la precisión y la imparcialidad es deseable en sí misma, y puede alcanzar un gran consenso en una sociedad crecientemente polarizada.

En conclusión, no deberíamos desaprovechar el gran activo con que cuentan nuestras administraciones e instituciones. Como en la era de los combustibles fósiles lo fue el petróleo, en la era de la informatización son los datos el recurso más abundante y preciado. Estemos a la altura de nuestro potencial.

Referencias bibliográficas

AEVAL (2010). *Fundamentos de Evaluación de Políticas Públicas*. Ministerio de Política Territorial y Administraciones Públicas. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia0.pdf#page=1

Brugarolas, P., De la Rica, S., & Gorjón, L. (2020). *El uso de datos administrativos para la investigación. El caso de Dinamarca: un ejemplo de buenas prácticas*. Fundación ISEAK. https://iseak.eu/wp-content/uploads/2020/11/El-sistema-danés-demicrodatos_DFG.pdf

De la Rica, S., Gorjón, L., & Lizarraga, I. (2020). *El contrato en prácticas en España: crónica de un desastre no anunciado*. Fundación ISEAK. https://iseak.eu/documentos/el-contratoen-practicas-en-espana-cronica-de-un-fracaso-no-anunciado-2/

Instituto para la Evaluación de Políticas Públicas (2020a). *Guía de Evaluación de Diseño de Políticas Públicas*. Ministerio de Política Territorial y Función Pública. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia_de_Evaluacion_de_Diseno.pdf#page=1

Instituto para la Evaluación de Políticas Públicas (2020b). *Guía de Evaluación de Implementación de Políticas Públicas*. Ministerio de Política Territorial y Función Pública. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia_de_Evaluacion_de_Implementacion.pdf#page=1

Instituto para la Evaluación de Políticas Públicas (2020c). *Guía de Evaluación de Resultados de Políticas Públicas*. Ministerio de Política Territorial y Función Pública. https://www.mptfp.gob.es/dam/es/portal/funcionpublica/evaluacion-politicas-publicas/Documentos/Metodologias/Guia_de_Evaluacion_de_Resultados.pdf#page=1

Instituto para la Evaluación de Políticas Públicas (2021). *Evaluación de políticas públicas*. Consultado el 3 de septiembre de 2021. https://www.mptfp.gob.es/portal/funcionpublica/evaluacion-politicas-publicas.html

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings*, *109*, 122-26. DOI: 10.1257/pandp.20191078

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

La gestión de los datos administrativos en España: diagnóstico y retos de futuro*

The Management of Administrative Data in Spain: Diagnosis and Challenges

Miguel Almunia
CUNEF Universidad

Pedro Rey-Biel
ESADE, Universidad Ramón Llull

Resumen

Los países de nuestro entorno han dado pasos importantes para gestionar y facilitar el acceso a la ingente cantidad de datos que generan y recopilan las Administraciones públicas, lo que supone una oportunidad para mejorar la evaluación de las políticas públicas, promover que el diseño de nuevas políticas se base en la evidencia disponible y expandir el conocimiento científico utilizando los registros administrativos. Sin embargo, a pesar de algunas iniciativas aisladas recientes, España se está quedando rezagada. A partir del análisis comparado de la experiencia en otros países avanzados y destacando las ventajas e inconvenientes de las distintas opciones que tenemos, planteamos propuestas de mejora en el acceso a datos administrativos en España con el fin de aprovechar el potencial de los datos existentes para impulsar las políticas basadas en la evidencia científica.

Palabras clave: *datos administrativos, acceso, investigación, políticas basadas en la evidencia.*
Códigos JEL: *B40, C81, C82, I38.*

Abstract

The neighboring countries have taken important steps to manage and facilitate access to the vast amount of data generated and compiled by public administrations, which represents an opportunity to improve the evaluation of public policies, promote the design of new policies building on the available evidence and expand scientific knowledge using administrative records. However, despite some recent isolated initiatives, Spain is lagging behind. Based on the comparative analysis of the experience in other advanced countries and highlighting the advantages and disadvantages of the different options that we have, we make proposals to improve the access to administrative data in order to take advantage of the potential of existing data to promote policies based on scientific evidence in Spain.

Keywords: *administrative data, access, research, evidence-based policies.*

* Una versión anterior de este artículo se publicó como Policy Brief de ESADE EcPol bajo el título «Por un cambio de cultura en la gestión de datos en España: una propuesta de reforma», disponible en: https://dobetter.esade.edu/es/informe-gestion-datos-covid

DOI: https://doi.org/10.32796/cice.2021.102.7313
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

La recolección y procesamiento de datos son actividades esenciales que los Estados deben realizar para cumplir muchas de sus funciones básicas. Por poner algunos ejemplos recientemente relevantes, las Administraciones públicas necesitan registros de la actividad económica de los contribuyentes para recaudar impuestos, registros de afiliación a la seguridad social para gestionar los sistemas de previsión social, o información sobre los contagiados (y sus contactos), ingresados y fallecidos para afrontar una pandemia. No es casualidad que el término «estadística» provenga originalmente del latín *statisticum*, «relacionado con el Estado».

Los enormes avances en las tecnologías de la información y la comunicación (TIC) permiten el almacenamiento de una cantidad cada vez mayor de datos a un coste cada vez más bajo. Esto ha llevado a los gobiernos de muchos países a mantener grandes registros administrativos sobre tributación, salud, educación y programas sociales. A su vez, la mejora de nuestra capacidad de procesamiento y análisis de datos supone una oportunidad de expandir el conocimiento utilizando los registros administrativos. La posibilidad de realizar evaluaciones de las políticas públicas y diseñar nuevas políticas basadas en la evidencia existente depende crucialmente de la disponibilidad y el aprovechamiento de los datos.

La administración realiza una evaluación interna de seguimiento de sus actividades mediante el uso de indicadores, que requieren del uso de datos administrativos, como por ejemplo en las políticas activas de empleo. Sin embargo, la administración no puede, ni debe, tener el monopolio del análisis de los datos que posee. No puede, en primer lugar, porque no es su función principal. La prioridad de la administración es gestionar las políticas públicas, no realizar un análisis profundo de los datos de los que dispone. En segundo lugar, porque los recursos públicos disponibles para este fin son reducidos. Aunque la administración cuenta con algunos organismos especializados en producir información estadística, como el Instituto Nacional de Estadística (INE) o el Centro de Investigaciones Sociológicas (CIS), los recursos económicos de los que disponen para evaluar políticas públicas son limitados. Esta limitada capacidad de análisis de los datos administrativos supone un despilfarro de los recursos destinados a recolectarlos y una ineficiencia respecto al enorme potencial que el uso de esos datos podría tener para el desarrollo y la evaluación de las políticas públicas. Además, no es deseable que la administración tenga el monopolio del análisis de estos datos, porque podrían surgir conflictos de interés al hacerlo. La creación de una entidad independiente como la Autoridad Independiente de Responsabilidad Fiscal (AIReF) ha significado un paso hacia delante en este sentido, que esperamos sea aún mayor con la reciente ampliación de su mandato actual, inicialmente restringido únicamente a la evaluación de políticas fiscales.

Permitir que la comunidad científica tenga acceso, de manera controlada, a registros administrativos resultaría en un enorme incremento de nuestra capacidad para analizarlos. Expandiría nuestro conocimiento sobre la efectividad de las políticas públicas, aumentaría la transparencia de la actuación de las administraciones y, por

lo tanto, profundizaría la calidad de nuestra democracia (Arellano, 2018). Los incentivos están alineados: la administración se beneficiaría porque su capacidad de analizar los datos es limitada, mientras que la comunidad investigadora tendría la oportunidad de explotar esta ingente fuente de información para realizar investigación académica de calidad. Los nuevos conocimientos obtenidos se podrían utilizar para diseñar políticas públicas basadas en la evidencia, algo que desafortunadamente es poco habitual en nuestro país.

Los datos administrativos presentan varias ventajas respecto a los datos provenientes de otras fuentes, como las encuestas, para la evaluación de las políticas públicas. Al cubrir a toda la población, se consigue un tamaño muestral mucho mayor, corrigiendo posibles sesgos de selección que a menudo afectan a las encuestas. Los datos administrativos tienen una tasa de no respuesta mucho menor, lo cual permite disponer de bases longitudinales (panel) más extensas en el tiempo y con menor desgaste *(«attrition»)*. Además, tienen menos errores de medición, aunque estos nunca se eliminan por completo. Por último, los registros administrativos se pueden cruzar con otras fuentes de datos utilizando identificadores individuales o de hogar. Esto es esencial para poder realizar análisis más profundos sobre causalidad y evaluación de políticas públicas.

Los datos administrativos también presentan algunas limitaciones que se deben tener en cuenta (Slemrod, 2016). En primer lugar, los datos se recogen atendiendo a objetivos de gestión, lo que afecta al formato, estructura y alcance de los mismos. Por ello, el número de variables disponibles en cada registro suele ser limitado. Por ejemplo, los datos fiscales no suelen tener información socioeconómica de los contribuyentes, más allá de la edad y el sexo. Esta es otra de las razones por las que la posibilidad de cruzar datos de distintos registros es esencial para explotar su verdadero potencial. Además, al haber sido recogidos con el fin principal de facilitar la gestión, la construcción de bases de datos para investigación a partir de registros administrativos supondrá un importante esfuerzo adicional. Idealmente, este trabajo lo deberían realizar conjuntamente funcionarios públicos e investigadores, dado que los primeros tienen el mejor conocimiento de los datos y los segundos tienen mayor experiencia analizando datos para investigación. Por último, el volumen de las bases de datos administrativos es grande, a veces difícil de manejar. Sin embargo, esta limitación es superable dado el rápido avance en las tecnologías de almacenamiento en la nube y el desarrollo de nuevas herramientas de análisis como el *machine learning* o la inteligencia artificial.

Un factor importante por considerar es la protección de la confidencialidad e integridad de los datos administraivos con información sensible sobre los ciudadanos. Para ello, deben utilizarse métodos de anonimización de los datos, eliminando identificadores individuales y garantizando que no se pueda trazar la identidad de ningún individuo concreto usando otras variables. Igualmente, deben crearse protocolos que garanticen que el acceso a los datos sea seguro desde este punto de vista, como han hecho otros países. Es importante resaltar que el anonimato no es incompatible con la creación de códigos individuales, que permitan identificar y emparejar variables

procedentes de distintas bases de datos pero pertenecientes a la misma unidad de observación (por ejemplo ciudadano, empresa o institución).

La mayoría de los países de nuestro entorno han creado instituciones y sistemas de acceso a los datos administrativos que protegen la confidencialidad. Algunos de estos sistemas se describen en detalle en la sección 3 de este artículo. Para diseñar sistemas similares en España no tenemos que reinventar la rueda, pero sí elegir entre las diversas opciones que se analizan en la sección 2 e inspirarnos en los modelos que han creado estos países y adaptarlos a nuestra estructura administrativa. La reciente Directiva UE 2019/1024 de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público da un claro soporte legal al respecto, puesto que prevé que «cualquier documento o conjunto de datos conservado por organismos públicos (incluidas empresas públicas y organismos de investigación financiados públicamente) sea puesto a disposición general del público, para fines comerciales o no comerciales, y en formatos abiertos, legibles por máquina, accesibles, fáciles de localizar y reutilizables, junto con sus metadatos». Esta Directiva debería haber sido objeto de transposición a la legislación nacional de los países miembros antes del 17 de julio de 2021, sin embargo la web de Transparencia del Gobierno apunta que «la actividad lleva cierto retraso y puede no llegar a trasponerse en el plazo previsto».

Afortunadamente, no partimos de cero. En España existen ya algunas bases de datos administrativos disponibles para la investigación, como la Muestra Continua de Vidas Laborales (MCVL) o el Panel de Declarantes del IRPF, que se describen en la sección 4. Además, existen casos aislados de investigadores que han conseguido obtener el acceso a ciertos datos, la mayor parte de las veces gracias a contactos personales y a eternas gestiones, no regidas con criterios objetivos sobre el interés público de las investigaciones que proponen. No existen, sin embargo, protocolos generalizados que regulen la colaboración y la cesión de datos entre las administraciones y la comunidad científica. Es obvio que en un país con un alto grado de descentralización, y con tantos niveles de administración como el nuestro, no es sencillo acordar y escoger entre las diversas formas de hacerlo. Por ello, precisamente, es necesario un debate sosegado, que se enriquezca con la experiencia ganada de otros países de nuestro entorno, con realidades políticas muy diversas.

La disponibilidad de mejores datos también permitiría diseñar medidas políticas bajo criterios que permitieran su evaluación, creando una colaboración activa, y no sólo pasiva, entre investigadores y administraciones. En este sentido, fuera de nuestras fronteras destacan iniciativas como el *Behavioral Insights Team* o las agencias del tipo *What Works (¿Qué Funciona?)* del Reino Unido. La creación de agencias que colaboren con investigadores para la realización y evaluación de políticas basadas en la evidencia puede ser uno de los muchos pasos necesarios hacia la modernización de la administración pública en España. Afortunadamente, muchos de estos pasos (disponibilidad de datos públicos, creación de una agencia de evaluación, apertura a la colaboración científica para el desarrollo de intervenciones que permitan basar

las políticas en la evidencia), están alineados, crean sinergias y pueden marcar un verdadero cambio en la forma de tomar decisiones políticas en nuestro país. Una señal esperanzadora en este sentido es que el Real Decreto-Ley por el que se aprobó el Ingreso Mínimo Vital incluye en su artículo 30 el compromiso de evaluar los efectos de esta política en concreto. Esperemos que esta evaluación no solo se lleve a cabo bajo los estándares metodológicos más exigentes, sino que siente el precedente para la evaluación de muchas otras políticas públicas.

2. Marco conceptual

Existen diversos modelos para facilitar el acceso público a datos administrativos, cada uno con sus ventajas y sus inconvenientes. Dichos modelos difieren fundamentalmente respecto a qué instituciones custodian los datos, cómo y a quién se facilita cada tipo de datos y en qué condiciones pueden obtenerse y tratarse. Naturalmente, la estructura administrativa de cada país es determinante para la elección de un modelo u otro. A continuación describimos las principales decisiones que se deben tomar en el diseño de un sistema de acceso a datos administrativos para investigación, destacando las ventajas e inconvenientes de las diferentes opciones. En la siguiente sección, describiremos algunos modelos concretos de países de nuestro entorno.

2.1. ¿Qué institución debe custodiar los datos?

En todos los países existen múltiples organismos públicos que recaban información de los ciudadanos y mantienen registros de datos. Una de las decisiones principales cuando se diseña un sistema de acceso a datos es qué institución (o instituciones) será la encargada de custodiarlos. Una opción es adoptar un modelo centralizado en el que solo una institución se encargue de esta tarea. Habitualmente este papel recae en el instituto nacional de estadística, que opera como un «tercero seguro»: recibe datos de diversas entidades gubernamentales –que no necesariamente quieren compartir datos entre sí– y se encarga de combinar los datos (usando identificadores individuales) y modificar su formato para que sean útiles para los distintos usuarios, como pueden ser los investigadores. Otra opción es adoptar un modelo descentralizado en el que cada institución generadora de datos se ocupa de la custodia y cesión de estos a los investigadores. También pueden existir modelos intermedios en los que varias instituciones se agrupen en consorcios para proveer acceso a sus datos, funcionando una de ellas como tercero seguro, y otras tengan un sistema de acceso aparte.

El modelo centralizado tiene dos ventajas principales. En primer lugar, la división de tareas entre instituciones que solamente proveen datos y un tercero seguro –que los custodia y cede– puede llevar a una especialización y ganancias de eficiencia.

Los aspectos legales de la cesión de datos a investigadores, así como la logística de los proyectos de investigación, requieren un proceso largo de aprendizaje y en estas actividades hay rendimientos de escala. En segundo lugar, un sistema centralizado proporciona una mayor transparencia al proceso desde la perspectiva de los investigadores, que saben inmediatamente a qué institución dirigirse si necesitan datos de un país concreto.

Por otro lado, los modelos descentralizados también tienen algunas ventajas. En primer lugar, determinadas instituciones pueden tener requisitos particulares de confidencialidad y por ello prefieren tener un control más directo sobre el acceso a sus propios datos. En segundo lugar, en países con una estructura gubernamental muy descentralizada, un sistema que replique esa estructura se puede ajustar mejor al diseño institucional.

2.2. ¿Cuáles deben ser los criterios de acceso a los datos?

La necesidad de proteger la confidencialidad de los datos administrativos merece una consideración especial a la hora de definir quiénes pueden tener acceso a ellos. Por otro lado, para asegurar que su uso tenga el máximo impacto posible, es deseable que se priorice el acceso de investigadores con proyectos bien definidos que tengan interés desde el punto de vista académico y/o desde la perspectiva de la evaluación de políticas públicas.

Para que el sistema funcione correctamente es importante que los criterios de acceso a los datos sean uniformes y transparentes, al contrario de lo que sucede actualmente en España. Cualquier investigador que desee acceder a los datos debería presentar una solicitud explicando el objeto de su estudio y especificando los registros a los que necesitaría acceder. Estas solicitudes deberían ser evaluadas por un comité científico en función de sus méritos científicos y el interés de la propuesta. Además, debería ser evaluada por un comité de productores de datos (representantes de las Administraciones públicas) para dictaminar la utilidad del estudio para el sector público y la viabilidad de ofrecer los datos requeridos en un formato apto para la investigación. Si estas evaluaciones tienen un resultado positivo, el investigador deberá firmar un contrato comprometiéndose a cumplir las reglas de uso y, especialmente, proteger la confidencialidad de los datos. Este contrato debe incluir la posibilidad de imponer sanciones al investigador si se incumplen las normas, con un objetivo disuasorio. Hay que tener en cuenta que cualquier tipo de infracción por parte de un investigador tendría también un efecto reputacional enorme, porque el resto de la comunidad científica podría ver en riesgo el acceso a los datos por la irresponsabilidad de una sola persona.

2.3. ¿El acceso a los datos debe ser de pago o gratuito?

Existen algunos argumentos para que el acceso a datos administrativos sea gratuito. En primer lugar, se trata de datos públicos que pertenecen al conjunto de la ciudadanía, y se podría esgrimir el principio de transparencia de los datos públicos para que se ofrezcan de manera gratuita. Además, la gratuidad evitaría cualquier tipo de discriminación del acceso entre investigadores con pedigrí, o que trabajan en instituciones con mayores recursos, e investigadores con menor experiencia o de instituciones más modestas.

Sin embargo, no se puede ignorar que la implementación de un buen sistema de acceso a datos como los que se describen en este documento tiene un coste que puede ser sustancial, ya que incluiría los salarios del personal que contrate el sector público (administrativos, expertos de sistemas informáticos, expertos en manejo de bases de datos, etc.) que sería adicional a las plantillas ya existentes. También habría que contabilizar el coste de los equipos informáticos y el software necesario para el funcionamiento del sistema. Aunque algunos de estos costes son fijos y solo se tienen que incurrir una vez, otros son variables y dependen directamente del número de proyectos de investigación que se desarrollen al mismo tiempo.

Por tanto, consideramos que sería razonable que se cobre algún tipo de tasa a los usuarios de este servicio que ayude a financiarlo. Una posible ventaja de este sistema es que los investigadores necesitarán obtener recursos para sus investigaciones en convocatorias competitivas, lo cual facilitará el filtrado de buenos proyectos y generará una selección positiva. Para que estas tasas no supongan una barrera a la entrada, se podría proporcionar acceso gratuito a muestras pequeñas de las distintas bases de datos para que los investigadores puedan explorarlos antes de tener que pagar las tasas. En este modelo, las tasas se cobrarían para el acceso a los datos completos, y especialmente cuando el uso de los datos suponga un mayor coste para la agencia proveedora del servicio. Por ejemplo, en los casos en que sea necesario hacer nuevos cruces de bases de datos o se requiera una gran cantidad de recursos computacionales. En la sección 3 se mencionan varios ejemplos de países en los que se cobran tasas a los investigadores por el uso de los datos.

2.4. ¿Deberían revisar los resultados de las investigaciones los organismos proveedores de datos?

Esta es una pregunta delicada porque afecta a la credibilidad de cualquier investigación que se realice con datos administrativos, así como a los incentivos que puedan tener los diferentes organismos para ceder sus datos. Por un lado, es razonable que los proveedores de datos tengan un interés en los resultados obtenidos y, además, quieran tener la posibilidad de revisar los resultados para comprobar que son factualmente correctos y que en ningún caso se viola la confidencialidad. Sin embargo, es fundamental que estos organismos se comprometan de antemano a

permitir la publicación de los estudios independientemente de si están de acuerdo o no con las conclusiones. En este sentido, un sistema de pre-registro de los estudios que se están llevando a cabo, de forma que posteriormente haya que reportar los resultados, podría contribuir a dar credibilidad al sistema y a evitar incentivos perversos[1].

2.5. ¿El acceso a los datos debería ser en salas seguras o a través de conexión remota?

En muchos casos, las instituciones proveedoras de datos prefieren que el acceso se realice exclusivamente en salas seguras de la propia institución, porque se considera que es la única manera de garantizar la seguridad del acceso y la confidencialidad.

Sin embargo, este modelo tiene muchas desventajas. En primer lugar, discrimina a investigadores no basados en la misma ciudad (a menudo la capital). En segundo lugar, en la práctica genera enormes pérdidas de eficiencia para los investigadores. Esto sucede porque los códigos que se utilizan para analizar grandes bases de datos a menudo tardan horas en ejecutarse y durante estos periodos de inactividad los investigadores no pueden conectarse a internet o trabajar en otras tareas, porque estas actividades no están permitidas en las salas seguras. Además, se debe tener en cuenta que los investigadores a menudo también tienen obligaciones docentes (y de gestión), lo que les impide pasar días completos en la sala segura trabajando en un único proyecto. Esto hace que los proyectos de investigación progresen a una velocidad mucho más lenta en los sistemas de acceso presencial a los datos.

Varios de los modelos que se presentan en la sección 3 demuestran que el acceso remoto a través de una conexión segura elimina estos costes de eficiencia a cambio de un incremento muy limitado en la probabilidad de que se quiebre la confidencialidad de los datos (Card *et al*., 2010; Almunia *et al*., 2019). Se trata, por lo tanto, de un aspecto que se debe tener muy en cuenta. En la sección 4 se menciona cómo algunas instituciones como el Banco de España ya se están moviendo en esta dirección.

3. Modelos de acceso a datos administrativos con fines de investigación

En esta sección se describen las características principales de tres modelos existentes de acceso a datos administrativos, incluyendo una descripción de algunos de los estudios más relevantes que se han realizado con cada modelo.

[1] Este sistema de pre-registro de estudios se ha instaurado con éxito en la investigación económica a través del *American Economic Association's registry for randomized controlled trials*, que se puede consultar en https://www.socialscienceregistry.org/

3.1. El modelo nórdico

Los países nórdicos (Dinamarca, Finlandia, Noruega y Suecia)[2] han sido pioneros en la creación de mecanismos para que la comunidad científica pueda realizar investigaciones utilizando sus datos administrativos de forma segura. En estos cuatro países, el organismo encargado de recopilar datos de diferentes registros administrativos es el instituto nacional de estadística respectivo. Las páginas web de estos institutos de estadística tienen un diseño moderno y accesible, y todas ellas incluyen una sección en la que se explica cómo se puede acceder a sus microdatos para fines de investigación.

Estos repositorios incluyen microdatos anonimizados sobre población (censos), empleo (historiales de cotización a la seguridad social), empresas (registros mercantiles), impuestos (renta, sociedades, IVA y especiales), sanidad (historiales hospitalarios), educación (resultados académicos), vivienda (precios de venta y alquiler), justicia (procesos judiciales), medio ambiente (datos históricos de clima y de partículas contaminantes) y comercio internacional (transacciones de importación y exportación).

Los criterios de acceso de investigadores a los datos varían entre países. En el caso de Dinamarca y Noruega es necesario tener una afiliación a una de las instituciones de investigación del país que esté aprobada por el instituto de estadística, mientras que los datos de Finlandia y Suecia se pueden obtener teniendo una afiliación con una institución dentro del Espacio Económico Europeo.

En todos los casos, los investigadores deben realizar una solicitud detallada en la que explican el diseño de su estudio y las bases de datos específicas a las que necesitan acceder para su investigación. En algunos casos los datos están disponibles en el formato deseado, pero en otros se requiere un trabajo por parte del instituto de estadística para vincular múltiples bases de datos. Existe también la posibilidad de que los investigadores aporten datos externos y el instituto de estadística los cruce con sus propios datos utilizando identificadores individuales, para crear una nueva base de datos anonimizada.

El acceso a los datos se realiza en todos los casos de forma remota por internet. Es decir, los investigadores reciben unas claves de acceso para conectarse a un servidor remoto donde están almacenados los datos y está instalado el software necesario para su análisis. Cada investigador tiene acceso exclusivamente a los datos de su estudio y por tanto no comparte el espacio con otros investigadores. El acceso remoto no permite la transferencia directa de archivos al ordenador de los investigadores. Cuando estos quieren extraer sus resultados, los archivos son revisados por personal del instituto de estadística para garantizar que se cumplen los requisitos de confidencialidad.

Los institutos de estadística cobran tasas por el uso de sus datos a los investigadores. Estas tasas incluyen un cargo por cada hora de trabajo que suponga la preparación de las bases de datos (alrededor de 100-150 euros por hora de trabajo) y también el coste de mantenimiento del servidor remoto (entre 150 y 300 euros por mes según el

[2] Habitualmente se incluye a Islandia en este grupo de países, pero no lo hemos incluido en nuestro análisis dado su reducido peso en términos de población.

volumen de datos almacenado), entre otros. Estas tasas garantizan que el sistema es sostenible, dado que asegura que se pueden financiar los costes, tanto de recursos humanos como de sistemas informáticos, aunque aumente el número de proyectos de investigación.

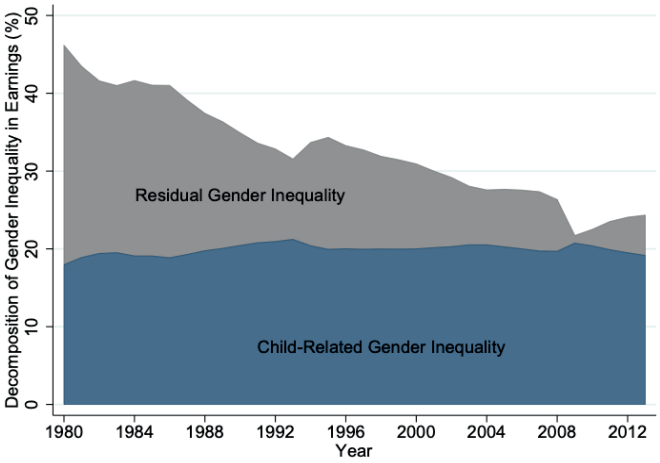
Ejemplos del Modelo Nórdico: Los efectos de distintos incentivos sobre el ahorro y los efectos de los hijos sobre la brecha salarial entre hombres y mujeres

a) Chetty *et al*. (2014a) estudia las decisiones de ahorro de los trabajadores en Dinamarca. Los investigadores combinan datos individuales de múltiples registros administrativos: declaraciones del impuesto sobre la renta, censo de población, una base de datos integrada de relaciones laborales, saldos de cuentas bancarias y aportaciones anuales a planes de pensiones (individuales y de empresa). Esta impresionante combinación de datos les permite observar todas las decisiones de ahorro del conjunto de la población danesa en el periodo 1995-2009. Los autores encuentran que los incentivos fiscales al ahorro previsional tienen un efecto moderado sobre el ahorro total (incluyendo planes de pensiones y otros instrumentos financieros). Sin embargo, estiman que un 85 % de los daneses son ahorradores «pasivos», que aceptan las contribuciones ofrecidas por sus empleadores por defecto, y solo un 15 % son ahorradores «activos», que reajustan su uso de distintos instrumentos de ahorro para aprovechar al máximo los incentivos fiscales, sin cambiar su ahorro total. Estos resultados tienen unas implicaciones claras para el diseño de los incentivos al ahorro previsional: es más efectivo regular que los planes de pensiones de empresa tengan, por defecto, un porcentaje alto de contribución que ofrecer un incentivo fiscal a todas las contribuciones a planes de pensiones.

b) Otro estudio que ha tenido un gran impacto es el de Kleven *et al*. (2019), que estudia, también para Dinamarca, la penalización laboral que sufren las mujeres cuando tienen su primer hijo. Los autores combinan datos de impuestos sobre la renta, censos de población, relaciones laborales y registros de nacimiento, que les permiten vincular a padres y madres con sus hijos. El principal hallazgo es que, con el nacimiento del primer hijo, la renta anual de la madre cae un 20 %, mientras que la del padre se mantiene constante. Esta brecha se mantiene inalterada durante los siguientes 10 años, y la misma tendencia se observa en las horas trabajadas, la tasa de empleo y en el salario por hora. Como se ve en la Figura 1, aunque la brecha salarial está decreciendo, el nacimiento del primer hijo explica cada vez un porcentaje mayor de la brecha en Dinamarca. Este estudio, junto con otros posteriores que se han hecho para otros países (incluyendo uno para España: De Quinto *et al*., 2020), indican que para cerrar la brecha salarial entre hombres y mujeres es fundamental diseñar políticas públicas que lleven a un reparto más equitativo de la crianza de los hijos entre padres y madres.

FIGURA 1
DESIGUALDAD DE GÉNERO EN RENTA DEBIDA A LOS HIJOS

A: Child-Related Inequality vs Non-Child Inequality



FUENTE: Kleven *et al*. (2019).

3.2. El modelo continental

Otros países de Europa continental han adoptado modelos diferentes al nórdico, pero que también están dando como resultado una fructífera colaboración entre la administración pública y la comunidad investigadora. La principal diferencia con el modelo nórdico es que no se construye alrededor del instituto nacional de estadística, sino que se han creado instituciones dedicadas específicamente a la tarea de recopilar y preparar los datos para su uso en investigación, además de la gestión de las solicitudes y los proyectos.

Además de los sistemas de acceso a datos en cada país, que describimos a continuación, en 2017 se creó INEXDA (International Network for Exchanging Experience on Statistical Handling of Granular Data), una red de bancos centrales (principalmente europeos, entre ellos el Banco de España) e institutos estadísticos. Los objetivos de este grupo son compartir experiencias en el manejo estadístico de datos granulares (también llamados microdatos) y homogeneizar los datos que gestionan los bancos centrales, de forma que las estadísticas producidas en distintos países sean comparables entre sí. El objetivo a largo plazo es que los datos estén disponibles para el análisis, el diseño de políticas públicas y la investigación. Aunque todavía no existe una lista completa de los datos que estarán disponibles, se pueden consultar las conclusiones del grupo de trabajo que ha definido los protocolos de acceso a los datos, las normas para combinar distintos registros y la información que se recolectará de los equipos encargados de cada proyecto de investigación.

3.2.1. Francia

En el caso de Francia, en el año 2010 se creó un centro de acceso seguro a datos (CASD) que actúa como tercero seguro y dispone de registros administrativos de impuestos, trabajo, empresas, finanzas, justicia, medio ambiente, agricultura y sanidad, todos ellos provenientes de los diferentes ministerios y del instituto de estadística francés (INSEE). El CASD está constituido como un «grupo de interés público» que reúne a varias instituciones: INSEE, GENES (las *grandes écoles* nationales), CNRS (centro nacional de investigación) y dos universidades: École Polytechnique y HEC Paris. Según se indica en su página web: «El objetivo principal de este grupo, de carácter industrial y comercial, es organizar e implementar servicios de acceso seguro a datos confidenciales para investigación, estudio, evaluación o innovación sin ánimo de lucro».

Una característica particular del CASD es que ha creado su propia tecnología, la *SD-box*, para permitir el acceso remoto a los datos por parte de investigadores externos. Se trata de un aparato (similar a un descodificador) que, conectado a un monitor y a una red de internet, permite a los investigadores tener acceso remoto a los datos de su proyecto. Al igual que el modelo nórdico, los investigadores interesados tienen que enviar una solicitud detallando el objetivo de su estudio y las bases de datos requeridas, y el servicio tiene un coste que depende de varios factores. En este caso, los investigadores con proyectos aprobados deben viajar a París para recibir un breve curso sobre las normas para proteger la confidencialidad de los datos.

Para su funcionamiento, el CASD se apoya en tres comités: un comité científico, un comité de productores de datos y un comité de seguimiento de la política de seguridad de los sistemas de información. El primero, formado por 16 investigadores, asesora a la dirección del CASD en materias de prospectiva, innovación, ética y estrategia científica. El segundo comité asesora a la dirección en asuntos relacionados con las condiciones de acceso a los datos, la documentación, el archivo y la difusión de información. El tercero asesora sobre temas relacionados con la protección de la confidencialidad e integridad de los datos. Una lista completa (e impresionante) de todos los estudios que se están realizando con datos del CASD está disponible.[3]

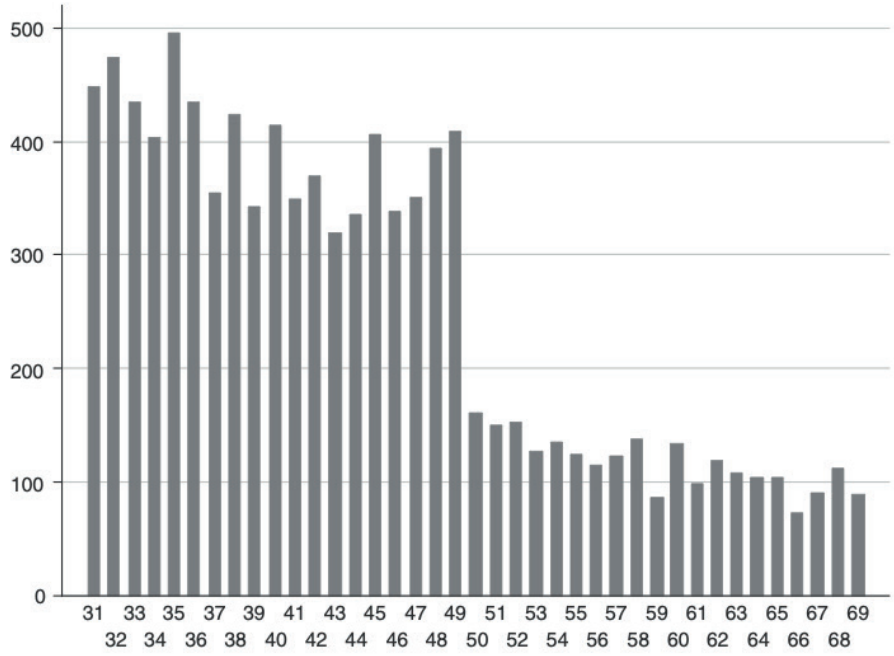
Uno de los grandes beneficiarios de la creación del CASD ha sido el Institut de Politiques Publiques, un think tank creado por la Paris School of Economics y las Grandes Écoles para promover la evaluación cuantitativa de las políticas públicas utilizando métodos de vanguardia. Este instituto ha tenido un notable impacto sobre el debate público en Francia, beneficiando al conjunto de la sociedad con sus evaluaciones independientes.

[3] https://www.casd.eu/en/projects-hosted-at-casd/

Ejemplo del Modelo Continental Francés. Evaluación de políticas públicas en Francia gracias a la creación del CASD en 2010

a) Garicano *et al*. (2016) analiza los efectos de las regulaciones laborales sobre el tamaño empresarial. Para ello utilizan datos sobre la población de empresas manufactureras en Francia, mostrando que hay una notable acumulación de empresas justo por debajo del umbral de 50 trabajadores, a partir del cual se aplican una serie de regulaciones que hacen el coste de los empleados mayor para las empresas. Los autores estiman que estas regulaciones reducen la competitividad global de las empresas francesas, al incentivar la actividad de las más pequeñas, que son menos productivas en promedio.

FIGURA 2
NÚMERO DE EMPRESAS POR TAMAÑO DEL EMPLEO EN FRANCIA



FUENTE: Garicano *et al*. (2016).

3.2.2. Alemania

La Agencia Federal de Empleo creó en el año 2004 un Centro de Datos para Investigación *(Forschungsdatenzentrum*, FDZ) que ofrece varias bases de datos provenientes de registros administrativos de la seguridad social alemana. Entre ellas

está la base integrada de historiales laborales (Integrated Employment Biographies, IEB), que contiene información sobre el censo de relaciones laborales del sector privado, incluyendo la remuneración y los días trabajados en cada empleo año a año, información sobre el nivel educativo y la ocupación del trabajador, el sector de la empresa, el régimen de dedicación (tiempo parcial o completo) y un identificador de establecimiento[4].

Existen tres modalidades de acceso a estos registros administrativos: presencial, remoto y a través de bases de datos reducidas para uso científico. Para facilitar el acceso en modo presencial a investigadores de todo el mundo, el FDZ ha llegado a acuerdos con diversas instituciones académicas y de investigación para establecer salas seguras de investigación en universidades de varios países como Estados Unidos (Michigan, Berkeley, Harvard, Princeton, Cornell, UCLA), Reino Unido (UCL, Essex), Canadá (Vancouver School of Economics), Francia (CASD) e incluso España (Universidad Carlos III de Madrid y Banco de España). El acceso a estos datos está abierto a investigadores de cualquier país del mundo y no tiene coste. Bajo estos acuerdos se están produciendo cientos de estudios anuales que cruzan la información proveniente de los distintos registros administrativos bajo condiciones de anonimato.

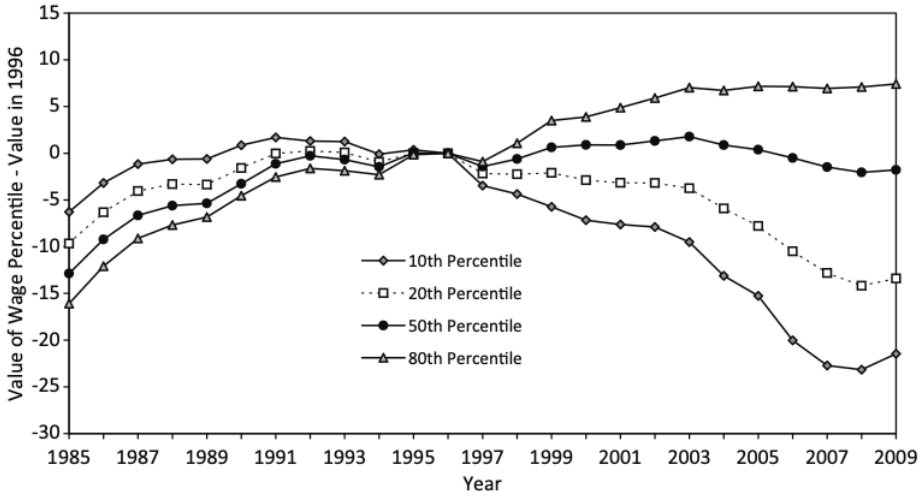
Ejemplos del Modelo Continental Alemán. Evaluación de la desigualdad salarial y de la carga del impuesto de sociedades entre empresarios y trabajadores.

a) Card *et al*. (2013) utilizan la base IEB para estudiar la desigualdad salarial. Los autores explotan la riqueza de los datos vinculados para estimar un modelo con efectos fijos de trabajador y establecimiento. Los resultados muestran que, a lo largo del periodo 1985-2009, se incrementaron tanto la dispersión en la productividad de los trabajadores como la dispersión en los salarios ofrecidos por las empresas (para un cierto nivel de productividad). En paralelo a estos procesos, creció la probabilidad de que los trabajadores más productivos se asociasen a los establecimientos con mayores salarios, dando como resultado un notable incremento en la desigualdad salarial en Alemania en este periodo. Esta investigación difícilmente hubiera podido realizarse con los datos de una encuesta o con una pequeña muestra de los datos administrativos, porque la identificación de los parámetros del modelo depende de los trabajadores que cambian de empresa: si ningún trabajador cambia de empresa, no se puede distinguir el efecto del trabajador y del establecimiento sobre el salario.

[4] Una empresa puede tener varios establecimientos.

FIGURA 3

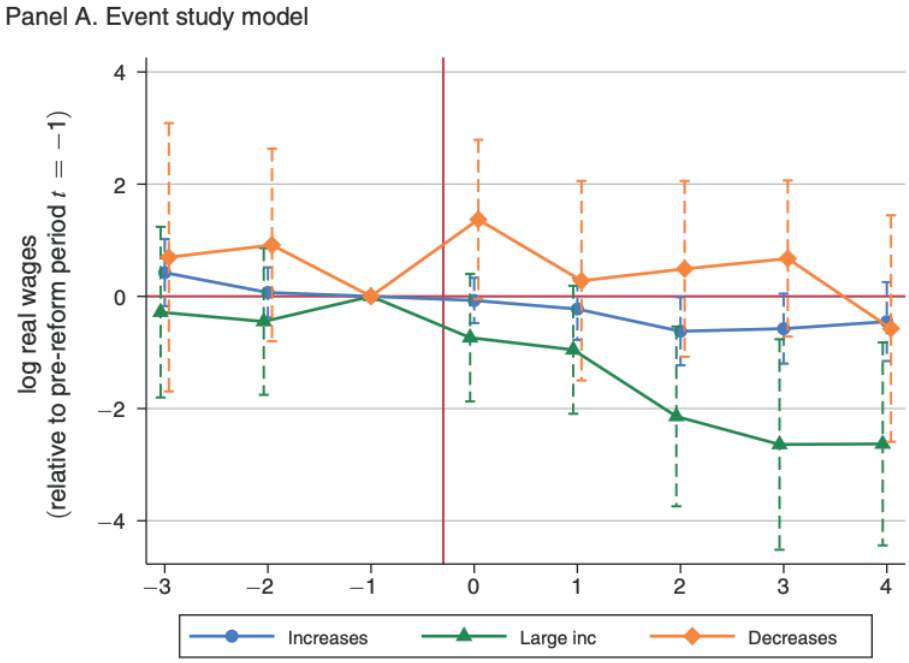
TENDENCIAS EN PORCENTAJES DEL LOGARITMO DE LOS SALARIOS DIARIOS PARA HOMBRES DE ALEMANIA DEL ESTE



FUENTE: Card *et al*. (2013).

b) Fuest *et al*. (2018) responden a una pregunta clásica: ¿quién soporta la carga de los impuestos sobre beneficios empresariales, el empresario o los trabajadores? La teoría nos dice que el sujeto que tiene la obligación legal de pagar un impuesto (en este caso, el empresario) no es necesariamente el que soporta toda la carga del impuesto, ya que puede ajustar su comportamiento de forma que parte de la carga se traslade a otro agente (en este caso, los trabajadores, a través de menores salarios). Los autores analizan miles de reformas de este impuesto local sobre los beneficios empresariales en Alemania para estimar hasta qué punto se traslada la carga impositiva a los trabajadores. Para ello, combinan datos a nivel municipal con el panel de empresas y trabajadores ofrecido por la seguridad social alemana, que incluye información sobre salarios a nivel individual. El principal resultado es que, cuando se produce una subida del impuesto sobre beneficios en un municipio, los trabajadores soportan el 51 % del coste de dicha reforma. Es importante destacar que no necesariamente bajan los salarios en términos nominales, sino que suben menos de lo que lo hacen en otros municipios comparables en los que no cambió la escala del impuesto.

FIGURA 4
LOGARITMO DE LOS SALARIOS REALES RELATIVOS ANTES DE LA REFORMA



FUENTE: Fuest *et al*. (2018)

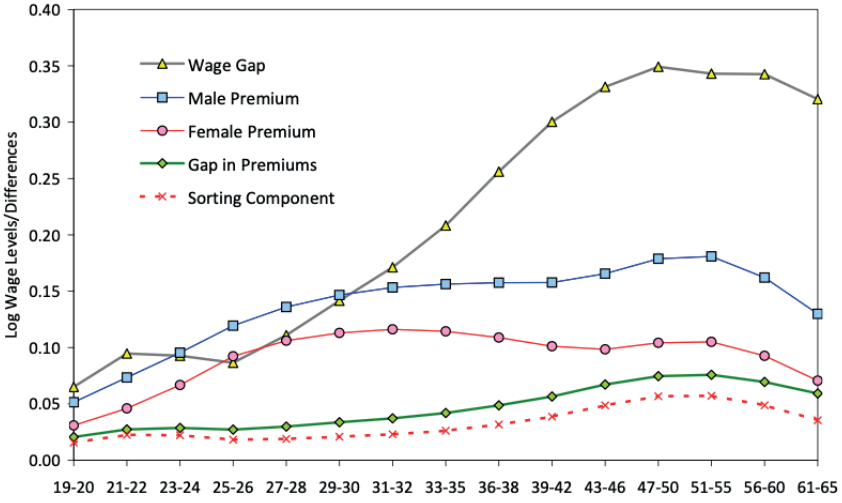
3.2.3. Portugal

Portugal ha desarrollado en los últimos años varios mecanismos de acceso a datos administrativos. Por un lado, en 2014 se formalizó la creación de un consorcio entre el *Instituto Nacional de Estatística* (homólogo del INE), la *Fundaçao para a Ciência e Tecnologia* (similar al CSIC) y la *Direção Geral das Estatística da Educação e Ciência*. El objetivo de este consorcio es proveer de acceso a los microdatos del INE a científicos afiliados a centros de investigación portugueses, incluyendo a estudiantes de doctorado y máster (en cuyo caso la solicitud debe ser refrendada por sus directores académicos). Los datos disponibles incluyen registros administrativos de empresas, relaciones laborales y comercio internacional, así como una gran variedad de encuestas con información demográfica, de salud y educación, entre otros. El acceso a los microdatos se debe realizar en las instalaciones del INE portugués o en centros seguros aprobados por el mismo.

Ejemplo del Modelo Continental Portugués. Las causas de la brecha salarial

a) Una de las bases de datos administrativos que más atención ha recibido de los investigadores son los *Quadros de Pessoal,* que recogen información detallada de todas las relaciones laborales en el sector privado (incluyendo duración de los contratos y remuneraciones) en Portugal durante varias décadas. Utilizando esta base de datos, Card *et al*. (2016) analizan la contribución de las empresas a la brecha salarial entre hombres y mujeres en Portugal. Gracias a la riqueza de estos datos, pueden distinguir dos efectos: por un lado, las mujeres tienen un menor poder de negociación salarial; por otro lado, las mujeres tienden a trabajar en empresas que pagan menores salarios. Estos dos efectos explican un 20 % de la brecha salarial total entre hombres y mujeres en Portugal.

FIGURA 5
EVOLUCIÓN DE LA BRECHA SALARIAL Y SUS COMPONENTES DURANTE EL CICLO VITAL



FUENTE: Card *et al*. (2016).

3.3. El modelo anglosajón

En los países anglosajones también ha habido un enorme interés recientemente por el uso de datos administrativos para la investigación. El modelo de acceso es generalmente más descentralizado que en los países nórdicos o de Europa continental, pues distintos organismos gubernamentales diseñan sus propios sistemas y protocolos de acceso a datos.

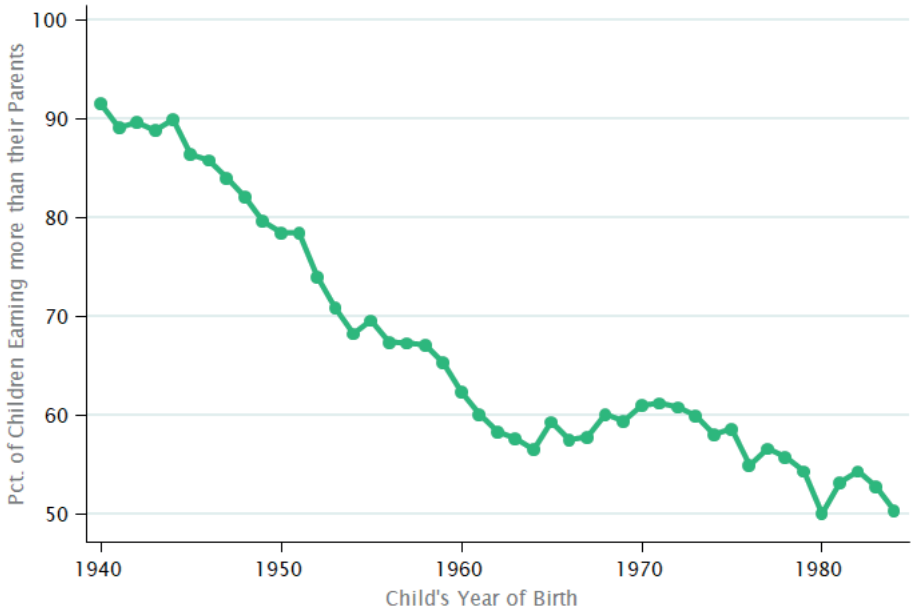
3.3.1. Estados Unidos

En Estados Unidos, un grupo de economistas de las universidades de Harvard y Berkeley escribió en septiembre de 2010 una carta abierta a la National Science Foundation (NSF) para pedir que se establecieran sistemas de acceso a datos administrativos (Card *et al*., 2010). En esta carta, argumentaban que se estaba realizando una gran cantidad de evaluaciones de políticas públicas en países europeos por la mayor disponibilidad de datos administrativos, y que Estados Unidos se estaba quedando atrás en este aspecto, perdiendo la posición de país líder en investigación en ciencias sociales. A raíz de esa iniciativa, instituciones como el Internal Revenue Service (la agencia tributaria estadounidense) pusieron en marcha sistemas de acceso a microdatos fiscales para investigación, dando lugar a una explosión de estudios sobre los efectos de las políticas públicas en Estados Unidos.

Ejemplo del Modelo Anglosajón en Estados Unidos. Nueva evidencia sobre movilidad intergeneracional usando los microdatos tributarios

a) Uno de los trabajos resultantes analiza los cambios en la movilidad intergeneracional en Estados Unidos (Chetty *et al*. 2014). Utilizando datos de declaraciones del impuesto sobre la renta, observan la renta de todos los estadounidenses nacidos en 1980-82 cuando tienen aproximadamente 30 años (en 2011-12) y vinculan esta información con la renta declarada por sus padres en el periodo 1996-2000, cuando los hijos eran adolescentes. Esto les permite estimar la relación entre la posición relativa de la renta de los padres y la de los hijos cuando son adultos. Los autores encuentran una gran variación en la movilidad intergeneracional entre distintas ciudades: la probabilidad de que un niño nacido en una familia que está en el 20 % más pobre de EE.UU. pase a estar en el 20 % más rico del país en su edad adulta es del 12,9 % en San José (en el *Silicon Valley* californiano) y solo del 4,4 % en Charlotte (en el estado sureño de Carolina del Norte). Comparando estos resultados con estudios de otros países, la movilidad intergeneracional en San José es similar a la de Dinamarca, pero la de Charlotte y otras ciudades del sureste es inferior a la de todos los demás países avanzados para los que existen datos. Los datos agregados de movilidad por ciudades se pueden consultar online. Estos resultados sugieren que EE.UU. ya no es, en gran parte de su territorio, una «tierra de oportunidades» y que la idea del sueño americano cada vez choca más con la realidad de una sociedad desigual en la que el éxito de una persona depende más de quiénes son sus padres que de su esfuerzo y mérito personal. Adicionalmente, el porcentaje de jóvenes que tiene mayor renta que sus padres al llegar a la edad adulta no ha parado de bajar (véase Figura 6). Este estudio, y otros relacionados que también hacen uso de datos tributarios, ha tenido un enorme impacto en el debate público sobre la desigualdad económica y racial en EE.UU.

FIGURA 6
PORCENTAJE DE HIJOS CON GANANCIAS MAYORES QUE SUS PADRES, POR AÑO DE NACIMIENTO



FUENTE: www.opportunityinsights.org/nationaltrends

3.3.2. Reino Unido

En el caso del Reino Unido, la agencia tributaria (Her Majesty's Revenue and Customs, HMRC) también ha creado un mecanismo para permitir el acceso de investigadores a datos fiscales y aduaneros, llamado HMRC DataLab. Al igual que en otros países, para obtener acceso a los datos se completa una solicitud en la que se explican los objetivos del estudio y las bases de datos necesarias. Para algunos proyectos se han combinado datos fiscales con otros datos externos, como por ejemplo los provenientes de registros mercantiles. El acceso a los datos solo se puede realizar presencialmente en una de las sedes de HMRC en Londres, lo cual limita parcialmente el acceso a investigadores residentes en otras ciudades del país. Además, el servicio no está sujeto al pago de ninguna tasa, una de las razones por las que no se ha realizado la inversión en un sistema de acceso remoto. El estudio de Almunia *et al*. (2019) realiza una comparación detallada del sistema británico con el finlandés.

Además, desde 2017 se puede acceder a microdatos del instituto de estadística británico (Office of National Statistics, ONS). El acceso se realiza a través de una red de centros seguros distribuidos por todo el país, donde los investigadores deben acudir en persona. Dada la restricción al acceso remoto, la pandemia ha supuesto una limitación en el acceso a datos para diversos proyectos de investigación.

Ejemplos del Modelo Anglosajón en el Reino Unido: Estudios sobre el efecto de las políticas fiscales en Reino Unido gracias al HMRC DataLab

a) Este sistema de acceso ha permitido que se realizaran investigaciones de alto impacto, como un estudio sobre los efectos de los impuestos a las transacciones inmobiliarias sobre el mercado de la vivienda (Best & Kleven, 2018), otro sobre los salarios obtenidos por estudiantes graduados en distintas universidades británicas (Belfield *et al*., 2018), una estimación del tipo impositivo efectivo que pagan las multinacionales extranjeras en Reino Unido en comparación con empresas locales del mismo tamaño (Bilicka, 2019) o el impacto de los incentivos fiscales a las donaciones a ONG y fundaciones (Almunia *et al*., 2020).

4. El acceso actual a datos administrativos en España

El acceso a datos administrativos para investigación en España es limitado en comparación con los países mencionados en la sección anterior, aunque no partimos de cero. España comparte con otros países europeos una gran ventaja a la hora de combinar diferentes registros administrativos: la existencia de un identificador individual que coincide con el número del documento nacional de identidad (DNI). Aunque para la gran mayoría de estudios, dicho identificador ha de ser anonimizado mediante un código, el identificador permitiría, en teoría, seguir a un mismo ciudadano cruzando distintas bases de datos de forma que se pudieran estudiar los efectos de variables pertenecientes a distintas fuentes.

Una visión de conjunto de la actual situación en España muestra a una serie de instituciones dando pequeños pero importantes pasos hacia el acceso a los datos. Sin embargo, muchas de estas iniciativas son parciales y aisladas, por lo que se las puede considerar como el sustrato sobre el que esperemos que vaya fermentando una visión más global que supere las muchas limitaciones actuales para acceder a los relativos escasos datos disponibles. En España existen tres fuentes básicas de datos administrativos a las que los investigadores han podido tener acceso: la Muestra continua de vidas laborales, el Panel de empresas-trabajadores y el Panel de declarantes del IRPF.

La Muestra Continua de Vidas Laborales (MCVL), disponible desde 2004, es una muestra aleatoria que contiene información (anonimizada) sobre un 4 % de los cotizantes y beneficiarios de la Seguridad Social. Se trata de una base de datos longitudinal que permite seguir el historial laboral de estos individuos a lo largo del tiempo. Estos datos se han utilizado en un gran número de estudios, por ejemplo para analizar la evolución de la desigualdad salarial a lo largo del ciclo económico (Bonhomme & Hospido, 2017) o evaluar la empleabilidad de los parados de larga duración (Bentolila *et al*., 2017).

El Panel de Empresas-Trabajadores (PET) se presentó en 2019 y supone un avance importante respecto a la MCVL. El muestreo se hace a nivel de empresa, en lugar

de trabajador, de forma que los investigadores observan a todos los individuos que durante el periodo 2013-2016 han estado afiliados al Régimen General de la Seguridad Social a través de cada empresa seleccionada en la muestra. La muestra cubre un 3-5 % de las empresas pequeñas (entre 1 y 9 trabajadores) y un 15% de las más grandes (más de 500 trabajadores). La manera de construir la muestra permitirá a los investigadores analizar cuestiones como los flujos de creación y destrucción de empleo, la contribución de los salarios a la desigualdad de la renta, los efectos del uso (y abuso) de los contratos temporales y la segregación horizontal y vertical de los trabajadores.

El Instituto de Estudios Fiscales (IEF) ofrece acceso a microdatos del Impuesto sobre la Renta de las Personas Físicas (IRPF) para el periodo 1999-2015 en dos formatos: muestras aleatorias anuales (que incluyen un 10-15 % del total de declaraciones anuales) y un panel de declarantes (que incluye un 3 % de las declaraciones anuales, en media). Las muestras son especialmente útiles para hacer ejercicios de microsimulación, mientras que el panel es útil para analizar los efectos de reformas fiscales sobre el comportamiento de los contribuyentes, como se hace en un estudio reciente que estima la elasticidad de la renta respecto a cambios en el tipo impositivo marginal (Almunia & López-Rodriguez, 2019).

Estas tres bases de datos administrativos se distinguen de las descritas anteriormente para otros países en que se trata de muestras representativas, pero que incluyen sólo un porcentaje de la población total. Desde el punto de vista del análisis estadístico, asumiendo que el muestreo esté hecho correctamente, es posible inferir características generales de la población a partir de una muestra. Sin embargo, muchas preguntas de investigación sólo se pueden abordar con datos poblacionales, porque las muestras dejan de ser representativas para subgrupos muy específicos. Por ejemplo, un estudio reciente de Llaneras *et al*. (2020) ha tratado de replicar el análisis de movilidad intergeneracional realizado por Chetty *et al*. (2014) para Estados Unidos, pero las limitaciones de los datos disponibles hacen que este análisis no sea directamente comparable (como ha argumentado Polavieja, 2020). Esta limitación también afecta a estudios sobre la desigualdad de la renta y la riqueza, en los que es crucial tener información detallada sobre las personas que están dentro del top 1 % de la distribución.

Otra desventaja de estas bases de datos es que, en la actualidad, no se pueden cruzar con otros registros administrativos. Por lo tanto, la MCVL no permite realizar un análisis completo de los efectos de cambios en el salario mínimo sobre distintos tipos de empresas, según su tamaño, sector o localización geográficas. Tampoco podemos estudiar con datos administrativos el efecto de distintos niveles de educación sobre la renta, o los efectos de tener hijos sobre la desigualdad salarial entre los cónyuges. Es paradójico que los investigadores españoles puedan acceder al universo de historiales laborales de Alemania a través de las salas seguras establecidas por el FDZ alemán en la Universidad Carlos III de Madrid y el Banco de España, pero no tengan acceso a los datos equivalentes de la seguridad social española. No obstante, la reciente apertura de salas seguras en distintas ciudades del Ministerio de

Seguridad Social, supone un primer paso en la buena dirección. Aunque a fecha de hoy no es posible disponer del universo de historias laborales, y los investigadores deben solicitar extracciones concretas de los datos que necesitan, se espera que en breve se elimine esta traba administrativa. La web de la Seguridad Social no tiene ninguna documentación sobre los datos que están disponibles en las nuevas salas seguras. Esta información debería publicarse (en español y en inglés) para que los potenciales investigadores sepan a qué datos pueden acceder. Quizá incluso se podría ofrecer una base de datos «ficticios» que mostrara valores ficticios pero que diera cuenta de las variables disponibles, como de hecho ya hace el FDZ alemán, de forma que se facilitase el desarrollo de preguntas posibles de investigación y el desarrollo de códigos para el análisis, antes de acudir a las salas seguras presencialmente a realizar la petición expresa de datos.

Una limitación adicional para el análisis es la segmentación geográfica. Por ejemplo, y respecto a las bases de datos tributarios, no incluyen a las Comunidades Autónomas de Régimen Foral (Navarra y País Vasco), por lo que no se puede obtener una visión del conjunto del país ni se pueden analizar los flujos migratorios (de personas y empresas) desde y hacia estas dos Comunidades Autónomas como respuesta a diferencias en la regulación.

Por otro lado, el Banco de España ha lanzado recientemente una iniciativa, el BE-Lab, para dar acceso a investigadores externos a varias de sus bases de datos. Desde julio de 2019, se puede acceder a la Central de Balances, que compila la documentación aportada anualmente por las empresas no financieras al Registro Mercantil. Un aspecto positivo de esta iniciativa es que una vez se ha creado la infraestructura para el acceso seguro a datos (en las dependencias del Banco de España), se han seguido dando pasos para facilitar el acceso. En este sentido, destaca el hecho que desde mayo de 2021, se puede acceder a los datos del BELab de forma remota, lo cual supone un avance muy positivo en línea con el punto 2.5 de nuestro marco conceptual.

Por su parte, el INE proporciona acceso a través de su página web a microdatos de las encuestas que realiza. Estas encuestas tienen información valiosa, pero tienen las desventajas propias de realizarse con muestras parciales de la población. Además, los únicos microdatos administrativos que se ofrecen en esta página son los de nacimientos y defunciones. Estos últimos son útiles para estadística descriptiva, pero, al no poder cruzarse con ninguna otra fuente de datos, no permiten realizar un análisis más profundo de relaciones entre distintas variables.

Finalmente, a nivel de las Comunidades Autónomas hay algunas iniciativas prometedoras, como el DataResSS (*Data for Research in Social Sciences*), un consorcio de la Barcelona Graduate School of Economics y el Instituto de Estadística de Cataluña (Idescat). El objetivo de este organismo es poner datos administrativos de Cataluña que posee Idescat a disposición de investigadores de vanguardia, tanto españoles como extranjeros. El DataResSS tiene un comité académico que evalúa las propuestas de investigación en función de su calidad académica, contribución al diseño de políticas públicas, cuestiones éticas relacionadas con la confidencialidad, y su factibilidad y extensión temporal esperada. Uno de los aspectos positivos del

diseño de DataResSS es la cooperación entre un grupo de universidades de máximo nivel con el instituto de estadística de la Comunidad Autónoma.

Quizá el paso reciente más esperanzador ha sido la reciente declaración conjunta del Instituto Nacional de Estadística (INE), la Agencia Tributaria (AEAT), el Ministerio de la Seguridad Social y el Banco de España comprometiéndose a «comenzar a trabajar conjuntamente en el desarrollo de un sistema de acceso a sus bases de datos con fines científicos de interés público». Dicha declaración supone un importante paso para desarrollar una estrategia más global por parte precisamente de las instituciones que ya han sido las impulsoras a nivel individual del relativamente escaso acceso a datos administrativos en España. No obstante, el comunicado no concreta ni el método de acceso a los datos que se busca ni la infraestructura que se utilizará, ni la estrategia que se seguirá para incorporar nuevas bases de datos. Nos parece además particularmente importante que se produzca un compromiso expreso para dotar de recursos, tanto informáticos como humanos, a esta iniciativa, de forma que no se quede en una intención por colaborar, sino que llegue a hacerse realidad, e inspire a pasos aún más ambiciosos. Para conseguirlo creemos que serían muy útiles las recomendaciones incluidas en el reciente informe de la AIREF, institución también relevante e implicada desde 2017 en la evaluación de políticas públicas y desde su formación en una estrategia de apertura de los datos administrativos.

5. ¿Cómo diseñar un sistema de acceso a datos administrativos para investigación en España?

En las secciones anteriores hemos explicado las distintas opciones que existen y hemos resumido la experiencia de países de nuestro entorno para demostrar que es posible ofrecer acceso a datos administrativos para investigación de forma segura y con importantes resultados en cuanto a la cantidad y calidad de la investigación sobre políticas públicas. A continuación, presentamos algunas reflexiones específicas sobre la mejor manera, a nuestro modo de ver, de adaptar estas experiencias a las particularidades de las administraciones públicas en España.

Nuestras reflexiones se centran en los cuatro aspectos clave que discutimos en la sección 2: ¿Qué institución, o instituciones, deben cumplir el papel de custodio de los datos y de «tercero seguro»? ¿Cuáles deberían ser los criterios de acceso a los datos? ¿Cómo se garantizará la confidencialidad de los datos? ¿Cómo se debería realizar el acceso a los datos?

5.1. ¿Qué institución debería jugar el papel de depositaria de los datos y tercero seguro?

Una posible ventaja de adoptar el modelo nórdico, donde el instituto de estadística centraliza el acceso a todos los registros administrativos, es que nuestro INE ya

dispone de un respaldo legal para recopilar y custodiar registros administrativos. Sin embargo, es importante ser conscientes de que el INE ya tiene asignadas una gran cantidad de funciones, como son la elaboración de la Contabilidad Nacional, el Censo de Población y numerosas encuestas. Por lo tanto, si se quiere asignar esta nueva tarea al INE, será indispensable dotarlo de los recursos humanos y técnicos suficientes para llevarla a cabo.

Una alternativa viable es seguir el modelo europeo continental, basado en la creación de consorcios de varias instituciones públicas con algunas entidades de investigación. En el caso de España, este tipo de consorcio podría incluir a las instituciones que ya han expresado formalmente su voluntad de colaborar: el INE, la Agencia Tributaria, la Seguridad Social, el Banco de España y la AIReF. Sería también deseable que las Comunidades Autónomas pudieran formar parte de este consorcio, quizá a través de sus propios institutos de estadística. De hecho, no es necesario que se forme un único consorcio. Podría comenzarse a trabajar con acuerdos puntuales para compartir ciertas bases de datos aprovechando que algunas instituciones, como el Banco de España, el INE o la AIReF, ya tienen creada una infraestructura de centros seguros. Paulatinamente se podrían ir incorporando otras instituciones una vez se compruebe que el sistema funciona adecuadamente y se comiencen a observar los resultados positivos que se obtienen de la evaluación de las políticas públicas.

5.2. Criterios de acceso a los datos

Los criterios de acceso a los datos deben ser uniformes y transparentes, al contrario de lo que sucede actualmente. Cualquier investigador que desee acceder a los datos debería presentar una solicitud explicando el objeto de su estudio y especificando los registros a los que necesitaría acceder. Esto implica que no se restrinja el acceso solamente a investigadores asociados a unas pocas instituciones de reconocido prestigio o a investigadores con un historial de publicaciones. El acceso debería estar abierto a cualquier investigador con una buena idea y, especialmente, a investigadores jóvenes (incluso estudiantes de doctorado, como argumentan Card *et al*. (2010) en el contexto de Estados Unidos). En este último caso se puede requerir el apoyo del director de tesis como garantía de compromiso con la protección de la confidencialidad y la calidad de la investigación.

Idealmente, deberían existir dos comités que revisen las solicitudes de acceso a los datos: el primero estaría formado por investigadores de reconocido prestigio que evalúe el interés de cada proyecto desde el punto de vista académico. El segundo estaría formado por personas de la institución generadora de los datos (o del «tercero seguro») que evalúen la viabilidad del proyecto desde el punto de vista logístico. Por ejemplo, si es posible combinar las bases de datos requeridas o acceder a ciertos datos que tengan una especial protección por su confidencialidad. Estos comités deberían reunirse con suficiente frecuencia, por ejemplo mensual o trimestralmente, para evaluar las solicitudes de acceso a los datos sin causar un retraso excesivo a los

proyectos de investigación. En cualquier caso, el proceso de solicitud debería ser sencillo: los procesos de investigación no siempre siguen procesos lineales, por lo que como punto de partida no se debería exigir más que una pregunta o hipótesis de investigación bien definida y una breve descripción del contexto en el que se plantea esa pregunta. Un formulario de pocas páginas es suficiente para recoger la información necesaria para la inmensa mayoría de proyectos de investigación. Un proceso demasiado extenso y complicado sería contraproducente ya que funcionaría como barrera a la entrada.

Por último, es crucial tener en cuenta la capacidad del sistema para absorber un cierto número de proyectos de investigación simultáneamente. Las limitaciones de capacidad pueden surgir por muchos factores: carga de trabajo en los comités de evaluación de proyectos, capacidad de las salas seguras (en caso de acceso presencial) o de los servidores (en caso de acceso remoto), retrasos en el proceso de revisión de nuevos resultados, etc.

5.3. Modo de acceso a los datos

El objetivo a medio plazo debe ser que exista la posibilidad de acceder a los datos de manera remota a través de internet, como se hace en varios países de nuestro entorno. Dicho esto, la secuencia habitual que han seguido otros países es desarrollar primero una infraestructura de centros seguros, en los que se maximiza el control de la actividad de los investigadores, que deben acudir presencialmente. Por ello, este suele ser el modo de acceso preferido de instituciones que ceden sus datos por primera vez. Sin embargo, la experiencia internacional indica que, una vez los sistemas están en marcha y funcionan adecuadamente, se suele producir un aumento de la demanda de nuevos proyectos que en poco tiempo supera la capacidad de estos centros seguros presenciales. Además, los centros seguros presenciales generan desigualdad territorial, pues el acceso es mucho más sencillo para los investigadores que residen en la misma ciudad (habitualmente la capital del país) que para los que están en otros lugares. Por todas estas razones, todos los modelos descritos anteriormente han terminado convergiendo hacia el acceso remoto a través de internet, haciendo uso de las tecnologías existentes. Como ha demostrado la experiencia reciente del INE, que cuenta con una sala segura para fines investigadores a la que desde hace poco se puede acceder de forma remota, es posible ofrecer acceso remoto y tener un control sobre la actividad de los investigadores muy similar al de los centros seguros presenciales. Por último, la opción del acceso remoto puede fomentar que accedan a los datos también investigadores ubicados fuera de España, expandiendo el grupo de investigadores potenciales que harán uso de los datos. En definitiva, la inversión a corto plazo en una tecnología segura de acceso remoto puede reducir enormemente los costes administrativos a largo plazo.

5.4. Cómo garantizar la confidencialidad

Este aspecto, que en muchas ocasiones se cita como una barrera insalvable, se ha solventado en todos los países de nuestro entorno cuyos sistemas hemos descrito anteriormente. ¿Existe alguna diferencia radical entre el sistema legal español y el de otros países europeos que impida el uso de datos administrativos con fines de investigación? La respuesta es no, por la sencilla razón de que si los datos están correctamente anonimizados y existe un control sobre el uso que se hace de ellos, en ningún momento se produce una quiebra de la confidencialidad. Tanto la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, como la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, son explícitas en este punto. La Directiva UE 2019/1024 de 20 de junio de 2019 aplica por igual a todos los países miembros de la Unión Europea e indica claramente que cualquier documento conservado por organismos públicos sea puesto a disposición general del público, por lo que, con las debidas precauciones, el impulso legislativo desde Europa hacia la accesibilidad de los datos es tajante.

El paso más básico en el proceso de anonimización de los datos consiste en sustituir los identificadores individuales (por ejemplo, DNI de personas o CIF de entidades jurídicas) por otros valores utilizando un algoritmo al que solo tenga acceso el depositario de los datos, pero nunca el investigador. En ocasiones, si las bases de datos administrativos son muy detalladas, puede ser necesario ir más allá y eliminar o distorsionar otra información (ubicación, edad, género, renta) que pueda permitir la identificación de un individuo o empresa concretos. Los detalles de estas técnicas van más allá del alcance de este artículo, pero es importante resaltar que se ponen en práctica en todos los países mencionados anteriormente y por lo tanto no hace falta inventar nuevos métodos.

La práctica habitual en los centros de acceso a datos administrativos es que los investigadores no pueden extraer ningún resultado hasta que no sea revisado por personal del centro donde los haya consultado. Este personal se cerciorará de que no se viola la confidencialidad, por ejemplo, poniendo como requisito que haya un mínimo de unidades (personas, empresas) en cada celda de las tablas de resultados.

6. Conclusiones

La modernización de la administración pública para poder implementar y evaluar políticas basadas en la evidencia pasa necesariamente por mejorar el acceso a los datos públicos y de diversos agentes facilitando la creación de conocimiento crucial tanto para el diseño y la evaluación de políticas como para la comunidad investigadora.

España se está quedando atrás respecto a otros países de nuestro entorno, que han creado agencias públicas de acceso a datos administrativos siguiendo distintos modelos que se adaptan a las características propias de su administración. El éxito de estas agencias nos permite usar como ejemplo los mejores modelos de cada una de

las importantes decisiones políticas que deben tomarse (centro seguro, confidencialidad, modo de acceso…) que mejor se adecúen a nuestra realidad territorial.

El cambio de cultura necesario para que nuestra muy fragmentada administración comparta y coordine sus datos supone un reto importante. Sin embargo, casos recientes muy notorios como la caótica gestión de la información durante la pandemia de COVID-19 abren una ventana de oportunidad para conseguir un consenso de todos los sectores de la sociedad sobre que este cambio de cultura es imprescindible.

Existen cada vez más casos aislados de éxito en nuestro país, e incluso iniciativas conjuntas de varias instituciones, sobre el potencial que abre el acceso público a los datos administrativos. Estos casos muestran que en España ni partimos de cero ni es inviable conseguirlo. La articulación de un Plan Nacional para impulsar la creación de una agencia de datos puede ir dando pequeños pasos, utilizando aquellos datos y administraciones menos sensibles, que afiancen el gran cambio necesario.

Referencias bibliográficas

Almunia, M., Harju, J., Kotakorpi, K., Tukiainen, J., & Verho, J. (2019). Expanding access to administrative data: the case of tax authorities in Finland and the UK. *International Tax and Public Finance*, *26*(3), 661-676. https://link.springer.com/article/10.1007/s10797-018-9525-0

Almunia, M., Guceri, I., Lockwood, B., & Scharf, K. (2020). More giving or more givers? The effects of tax incentives on charitable donations in the UK. *Journal of Public Economics*, *183*, 104114. https://www.sciencedirect.com/science/article/abs/pii/S0047272719301768

Arellano, Manuel (2018, 20-21 de agosto). *El acceso a los microdatos administrativos públicos: la nueva frontera de la investigación económica y social* [sesión de conferencia]. XVIII Aula de Verano Ortega y Gasset, Universidad Internacional Menéndez Pelayo (UIMP). http://www.cemfi.es/~arellano/arellano-presentacion-uimp-2018.pdf

Belfield, C., Britton J., Buscha F., Dearden L., Dickson M., Van Der Erve L., Sibieta L., Vignoles A., Walker I., & Yu Zhu (2018). *The relative labour market returns to different degrees*. Institute of Fiscal Studies. https://www.ifs.org.uk/publications/13036

Bentolila, S., García-Pérez, J. I., & Jansen, M. (2017). Are the Spanish long-term unemployed unemployable? *SERIEs*, *8*(1), 1-41. https://link.springer.com/article/10.1007/s13209-017-0155-z

Best, M. C., & Kleven, H. J. (2018). Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK. *The Review of Economic Studies*, *85*(1), 157-193. https://www.henrikkleven.com/uploads/3/7/3/1/37310663/bestkleven_landnotches_sep2016.pdf

Bilicka, K. A. (2019). Comparing UK tax returns of foreign multinationals to matched domestic firms. *American Economic Review*, *109*(8), 2921-53. https://www.aeaweb.org/articles?id=10.1257/aer.20180496

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, *131*(2), 633-686. https://doi.org/10.1093/qje/qjv038

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly journal of economics*, *128*(3), 967-1015. https://doi:10.1093/qje/qjt006

Card, D., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *American economic association, ten years and beyond: Economists answer NSF's call for long-term research agendas*. https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf

Chetty, R., Friedman, J. N., Leth-Petersen, S., Nielsen, T. H., & Olsen, T. (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark. *The Quarterly Journal of Economics*, *129*(3), 1141-1219. http://www.rajchetty.com/chettyfiles/ret_savings.htm

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553-1623.

https://eml.berkeley.edu/~saez/chetty-friedman-kline-saezQJE14mobility.pdf

De Quinto, A., Hospido L., & Sanz C. (2020). *The Child Penalty in Spain*. Documento Ocasional 2017, Banco de España. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/20/Files/do2017e.pdf

Directiva (UE) 2019/1024 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público (versión refundida). *Diario Oficial de la Unión Europea*. http://data.europa.eu/eli/dir/2019/1024/oj

Fuest, C., Peichl, A., & Siegloch, S. (2018). Do higher corporate taxes reduce wages? Micro evidence from Germany. *American Economic Review*, *108*(2), 393-418. https://www.aeaweb.org/articles?id=10.1257/aer.20130570

Garicano, L., Lelarge, C., & Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review*, *106*(11), 3439-79. https://www.aeaweb.org/articles/pdf/doi/10.1257/aer.20130232

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181-209. https://www.aeaweb.org/articles?id=10.1257/app.20180010

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings*, *109*, 122-26. https://www.aeaweb.org/articles?id=10.1257/pandp.20191078

Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. *Boletín Oficial del Estado*, nº 276, de 17 de noviembre de 2007. https://www.boe.es/eli/es/l/2007/11/16/37

Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno. *Boletín Oficial del Estado*, nº 295, de 10 de diciembre de 2014. https://www.boe.es/eli/es/l/2013/12/09/19/con

Llaneras, K., Medina O., & Costas E. (2020). *Atlas de Oportunidades*, proyecto conjunto de la Fundación COTEC y la Fundación Felipe González. https://www.cotec.es/fundacionfelipegonzalez/oportunidades/proyecto/

Polavieja, J. G. (2020). Grandes Datos, Grandes Sesgos, Grandes Errores: Sobre el Atlas de Oportunidades. *Revista Internacional de Sociología*, *78*(3), 166. https://www.javierpolavieja.com/sobre-el-atlas-de-oportunidades

Portal de transparencia. Gobierno Abierto. Compromiso 2. *Incorporación de Directiva (UE) 2019/1024*. Consultado el 16 de octubre de 2021. https://transparencia.gob.es/transparencia/transparencia_Home/index/Gobierno-abierto/seguimientoIVPlanGA/seguimiento_C2/2-3-1-IncorporacionDirectiva.html

Real Decreto-ley 20/2020, de 29 de mayo, por el que se establece el ingreso mínimo vital. *Boletín Oficial del Estado*, n.º 154, 1 de junio de 2020. https://www.boe.es/eli/es/rdl/2020/05/29/20/con

Slemrod, J. (2016). Caveats to the research use of tax-return administrative data. *National Tax Journal*, *69*(4), 1003-1020. http://www.ntanet.org/NTJ/69/4/ntj-v69n04p1003-1020-caveats-research-use-tax-return-data.pdf?v=%CE%B1&r=19426893815398216

CUADERNOS ECONÓMICOS ICE
INFORMACIÓN COMERCIAL ESPAÑOLA

La evaluación de impacto de las políticas activas de empleo en España: el deseo y la realidad

Impact evaluation of active employment policies in Spain: desire and reality

Inmaculada Cebrián López
Universidad de Alcalá

Begoña Cueto Iglesias
Francisco Javier Mato Díaz
Universidad de Oviedo

Resumen

Este artículo revisa aspectos clave relacionados con la evaluación del impacto de las políticas activas de empleo en España. Constatando que los Planes Anuales de Política de Empleo contemplan la evaluación, se advierten diversas insuficiencias para captar los efectos causales de las políticas desarrolladas. Frente a las dificultades de homogeneización de políticas e indicadores, se señala que el modelo descentralizado del Fondo Social Europeo ha potenciado las evaluaciones de impacto con enfoque contrafactual durante los últimos años. Finalmente, se analizan aspectos relacionados con los datos administrativos y las fuentes de información que podrían facilitar las evaluaciones de impacto en España.

Palabras clave: *políticas activas de empleo, evaluación de impacto, contrafactual, Fondo Social Europeo, datos administrativos.*

Clasificación JEL: *J38, J68.*

Abstract

This article reviews key aspects related to evaluating the impact of active employment policies in Spain. Noting that the Annual Employment Policy Plans contemplate evaluation, various shortcomings regarding the estimation of the causal effects of the policies are raised. As compared with the difficulties faced by homogenization of policies and indicators, it is pointed out that the decentralized model of the European Social Fund has promoted impact evaluations with a counterfactual approach in recent years. Finally, aspects related to administrative data and sources of information that could facilitate impact evaluations in Spain are analyzed.

Keywords: *active employment policies, impact evaluation, counterfactual evaluation, European Social Fund, administrative data.*

DOI: https://doi.org/10.32796/cice.2021.102.7314
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

En 1997, con la transferencia de las políticas activas de empleo a las comunidades autónomas, se pone en marcha la primera de las reformas laborales que centra su objetivo principal en la flexibilización del mercado de trabajo al tiempo que se fomenta el empleo indefinido como garantía de estabilidad laboral. Ni esta ni ninguna de las posteriores (2001, 2006, 2010, 2012) modificó sustancialmente el modelo de políticas activas, a pesar de que en todas se fueron modificando algunos elementos que tenían mucho que ver con las políticas activas de empleo, por ejemplo, el programa de incentivos económicos a la contratación, los contratos de formación y aprendizaje, las empresas de trabajo temporal (ETT) como intermediarios laborales, etc.

En España, la promoción y desarrollo de las políticas activas de empleo (PAE) tiene lugar dentro del Sistema Nacional de Empleo (SNE). Está integrado por el Servicio Público de Empleo Estatal (SEPE) y los 17 servicios públicos de empleo autonómicos (SPE). El SEPE y los SPE financian, gestionan y ejecutan las PAE, siendo el SEPE el coordinador.

El Sistema de Información de los Servicios Públicos de Empleo, conocido como SISPE, es la principal fuente de información, comunicación y coordinación del SNE. Por otro lado, el principal instrumento de colaboración de todos los SPE es la Conferencia Sectorial de Empleo y Asuntos Laborales, en la cual se determinan los servicios y programas que se van a desarrollar a nivel comunitario con cargo a los fondos distribuidos por el SEPE.

La financiación del SNE consta de cuatro niveles administrativos: la Unión Europea, la Administración General del Estado a través del SEPE, las comunidades autónomas y las entidades locales.

La Unión Europea es un agente financiador de políticas activas de empleo, principalmente a través del Fondo Social Europeo (FSE), Programas Operativos plurirregionales, regionales y de la Iniciativa de Empleo Juvenil. El SEPE desarrolla su papel como agente financiador, gestor, ejecutor y coordinador en el ámbito de las políticas activas de empleo.

Las comunidades autónomas son agentes financiadores, gestores y ejecutores de políticas activas de empleo, a través de sus respectivos SPE. Al tener las competencias en el ámbito de las políticas activas de empleo, se posicionan como agentes principales en la ejecución de esas políticas. Por otro lado, ejercen un papel significativo en su financiación, tanto en lo relativo a la financiación vinculada a programas y servicios integrados en el Plan Anual de Política de Empleo (PAPE) como en lo que respecta a la financiación de otras acciones propias no integradas en el PAPE o por la financiación que reciben de manera directa de la Unión Europea.

A nivel local, la gran mayoría de entidades locales (diputaciones, municipios, cabildos y consejos, comarcas, áreas metropolitanas y mancomunidades) juegan un papel como agentes financiadores, gestores y ejecutores de políticas activas de empleo a nivel de sus respectivos territorios.

La evaluación de las políticas activas realizada por la Autoridad Independiente de Responsabilidad Fiscal (AIReF) en 2019 puso de manifiesto las dificultades de coordinación en el Sistema Nacional de Empleo, lo que obstaculiza establecer objetivos y planificar. Así, en el sexto eje de objetivos estructurales del PAPE 2020, se indica que uno de los objetivos es la mejora de la gestión, coordinación y colaboración en el seno del Sistema Nacional de Empleo. Esta mejora contribuiría a establecer una evaluación rigurosa del gasto y del impacto de las medidas aplicadas, contribuyendo a la puesta en marcha de procedimientos que sean eficientes para destinar los fondos a las prácticas más eficaces para lograr los objetivos de las PAE.

En los apartados siguientes, se presentan una serie de análisis y reflexiones sobre el marco que regula las PAE y su evaluación, el FSE y su evaluación, así como la posibilidad que brindan los datos registrados en la Administración pública para proceder a la evaluación de impacto. Finalmente, se presentan algunas propuestas para una mejor gestión de las PAE y de su evaluación, y para favorecer la implementación de las prácticas más adecuadas.

2. La evaluación en el marco legislativo

La evaluación de las políticas activas está prevista formalmente en la mayor parte de la legislación. El artículo 39 del Real Decreto Legislativo 3/2015 por el que se aprueba el texto refundido de la Ley de Empleo establece que «el Gobierno presentará una memoria sobre el gasto, incluido el necesario para la gestión del sistema, y los resultados de las políticas activas de empleo en el conjunto del Estado».

En la Estrategia Española de Activación para el Empleo 2017-2020, se concede «especial relevancia a la planificación, evaluación y seguimiento de las políticas de activación para el empleo en base a resultados, una de las líneas de actuación prioritarias expresamente acordadas en la Conferencia de Presidentes del 17 de enero de 2017». En la misma, se establecen dos niveles para la evaluación basados en resultados:

1. Cuatro evaluaciones distintas, a realizar de forma anual o bienal:
   a) Una evaluación estratégica del cumplimiento de objetivos.
   b) Una autoevaluación anual del desempeño.
   c) Una evaluación externa bienal del desempeño junto con la Red Europea de Servicios Públicos de Empleo.
   d) Una evaluación anual de la calidad, impacto, eficacia y eficiencia del conjunto del sistema de formación profesional para el empleo.
2. Evaluaciones a realizar en cada uno de los servicios y programas concretos.

Uno de los objetivos estratégicos, el número 6.6, es el «Desarrollo e implantación de una metodología común para la evaluación del impacto de los programas de políticas de activación». De esta forma, se establecía que durante la vigencia de la Estrategia se acordaría e implantaría «una metodología para la evaluación de los

programas y medidas concretos recogidos en el PAPE». También se menciona una metodología sobre la recogida de datos y la selección de contrafactuales[1]. En la misma normativa, se establece que los «resultados de todas las evaluaciones se harán públicos, en cumplimiento del principio de transparencia y como factor determinante para la modernización y mejora del Sistema Nacional de Empleo».

La cuestión es hasta qué punto se ha avanzado en la consecución de estos objetivos. Ciertamente, se han cumplido parte de los objetivos al haber encargado a la AIReF la evaluación del conjunto de políticas activas (AIReF 2019; 2020). Así, en junio de 2019 se publicaba el tercer Estudio del *Spending Review* sobre Políticas Activas de Empleo[2] y, en octubre de 2020, la segunda fase del *Spending Review* sobre Incentivos a la contratación y al trabajo autónomo[3], sobre las que hablaremos más adelante.

En la elaboración de los Planes Anuales de Política de Empleo, PAPE, también se destaca la evaluación del cumplimiento de objetivos. El PAPE, que se publica anualmente, establece los objetivos a alcanzar en cada año en términos de empleo y los indicadores que se utilizarán para valorar el grado de consecución. En este sentido, se indica que este Plan se configura como «instrumento de evaluación de los servicios y programas de políticas activas de empleo desarrolladas por el Servicio Público de Empleo Estatal y las comunidades autónomas».

A partir de esta breve revisión, podemos concluir que la evaluación de impacto ya está incluida en el marco legislativo de las políticas de empleo. Sin embargo, no parece que, efectivamente, se hayan puesto en marcha varias de las medidas consideradas. O, si se han puesto en marcha, no parece que sus resultados hayan sido públicos, con la excepción de las evaluaciones realizadas por la AIReF.

2.1. Indicadores en el PAPE

En cada uno de los PAPE de los últimos años, se han establecido distintos indicadores, con sus respectivos componentes, con el fin de valorar la consecución de los distintos objetivos clave, estratégicos y estructurales.

El PAPE 2019[4] estableció un conjunto de 22 indicadores formados por 27 componentes. Esto supuso una simplificación importante respecto a años previos, aunque

[1] El enfoque contrafactual en la evaluación de políticas implica la existencia de un grupo de control que sirva de comparación para saber qué hubiera pasado al grupo de tratamiento, si éste no hubiera participado en la política a evaluar. Como se comentará más adelante, una de las claves en este tipo de evaluaciones es seleccionar un grupo de control adecuado que evite la existencia de sesgos de selección.

[2] https://www.airef.es/es/politicas-activas/

[3] https://www.airef.es/es/spending-review-fase-ii-estudio-incentivos-a-la-contratacion-y-al-trabajo-autonomo/

[4] Resolución de 12 de marzo de 2019, de la Secretaría de Estado de Empleo, por la que se publica el Acuerdo del Consejo de Ministros de 8 de marzo de 2019, por el que se aprueba el Plan Anual de Política de Empleo para 2019, según lo establecido en el artículo 11.2 del texto refundido de la Ley de Empleo, aprobado por el Real Decreto Legislativo 3/2015, de 23 de octubre (BOE núm. 64, de 15 de marzo de 2019).

seguían siendo muchos y muy complejos en su cálculo. En el caso del PAPE 2020[5], se han establecido 8 indicadores para los objetivos clave, 17 para los estratégicos y 33 para los estructurales. En este caso, hay que tener en cuenta que varios de ellos se han establecido durante la situación de pandemia y su objetivo era, fundamentalmente, monitorizar en qué medida se recupera la situación de febrero de 2020.

Una de las debilidades de estos indicadores para ser utilizados como instrumento en la evaluación de las políticas de empleo es que no miden la eficacia ni la eficiencia de los distintos programas de empleo, sino que están influidos, principalmente, por la coyuntura económica del territorio en que se implementan estos programas[6]. Por ejemplo, en los indicadores de los objetivos clave se calcula: la transición del desempleo al empleo, en proporción al número de desempleados inscritos en los SPE; el número de personas que dejan de estar inscritas como desempleadas en los SPE, en proporción a los desempleados inscritos; las transiciones al empleo dentro de los 6 (12) meses siguientes a su inscripción en los SPE.

Estos indicadores tienen como objetivo medir la *contribución* a la reducción del desempleo y a su duración. Sin embargo, las cifras que se obtengan van a estar más determinadas por la estructura económica de la zona en que se llevan a cabo las políticas, y por su coyuntura, que por la eficacia de las políticas.

La mayor parte de los indicadores incluidos miden la cobertura de las medidas puestas en marcha. Así, se hace referencia a los demandantes a quienes se les proporcione atención, respecto al total de demandantes, quienes hayan recibido un itinerario, algún tipo de formación u oportunidades de empleo.

En 2019, había dos indicadores que tenían como objetivo, específicamente, evaluar el impacto de los programas. El primero trataba de medir la cobertura y el impacto que había tenido la formación profesional para el empleo impartida en los demandantes de empleo. El segundo planteaba medir la cobertura y el resultado de la atención prestada para el fomento del empleo autónomo y el emprendimiento y la eficacia en términos de inserción en el Régimen Especial de Trabajadores Autónomos (RETA).

Aunque en estos dos indicadores se establece una comparación entre los demandantes que participaron en programas de empleo con aquellos que no lo hicieron, no tienen en cuenta las características de las personas beneficiarias ni los procesos de selección que llevaron a que unos demandantes sí participaran en la política y otros no. Como consecuencia, son indicadores sesgados. Existe, además, la posibilidad de que, dado que la financiación está condicionada a estos indicadores, se opte por

[5] Resolución de 26 de octubre de 2020, de la Secretaría de Estado de Empleo y Economía Social, por la que se publica el Acuerdo de Consejo de Ministros de 20 de octubre de 2020, por el que se aprueba el Plan Anual de Política de Empleo para 2020, según lo establecido en el artículo 11.2 del texto refundido de la Ley de Empleo, aprobado por el Real Decreto Legislativo 3/2015, de 23 de octubre (BOE núm. 289, de 2 de noviembre de 2020).

[6] Como muestra del equívoco que pueden causar estos errores, el trabajo de Vila *et al*. (2020) señala que la evaluación contrafactual reduce en cerca de dos tercios el efecto aparente de los programas de formación y de inserción de la Comunidad Valenciana, mientras que eleva el efecto de los programas de fomento del empleo.

seleccionar los participantes que pueden tener mejor inserción laboral (independientemente de que la política les resulte útil o no, dando lugar al efecto *creaming*) o por llevar a cabo políticas en los que sea más fácil obtener un buen indicador. Por ejemplo, llevando a cabo cursos de corta duración en lugar de otros más largos; o dirigiendo las políticas a colectivos con menos dificultades, porque será más fácil obtener mayores tasas de transición al empleo que si se dirigen a grupos con graves problemas de empleabilidad.

Estos indicadores no se han incluido en 2020. En el PAPE de dicho año se indica que «el diseño del sistema de indicadores no puede ni debe limitarse a la información cuantitativa basada en los datos de la gestión de cada Servicio Público de Empleo que ofrece solamente una perspectiva de impacto de las medidas en el correspondiente ámbito territorial». Por ello, se habla de un «nuevo enfoque de evaluación» de manera que se lleve a cabo un «análisis integral de la "política de activación"» para conocer hasta qué punto se están alcanzando sus objetivos y cuáles son las medidas que se demuestran más eficientes para ello.

Este párrafo muestra que no se entiende lo que significa evaluar el impacto de un programa de empleo. Una evaluación de impacto debe aportar información que muestre en qué medida los cambios en la situación (en este caso, laboral) de las personas pueden atribuirse a la participación en un programa. La clave, por tanto, está en establecer una relación causal entre el programa y un resultado de interés.

Conocer cuántas personas se han formado o a cuántas se les ha ofrecido un itinerario de inserción es relevante. También, saber en qué medida las personas registradas en los servicios públicos de empleo transitan a la ocupación. Pero estos resultados no implican que haya impacto. Tras haber pasado por una acción formativa o por un itinerario se puede haber obtenido un empleo, pero ello no implica que *se deba a* que se ha participado en dichas políticas. Por eso son necesarias las evaluaciones de impacto, porque es necesario saber en qué medida las políticas contribuyen, en sí mismas, a que los participantes accedan al empleo. No deben ser el único factor a tener en cuenta en el diseño de las políticas, pero sí deberían considerarse.

Los PAPE se centran en indicadores de seguimiento y de resultados, pero no de impacto. Como se verá más adelante, sí que se han llevado a cabo evaluaciones de impacto, en el ámbito académico[7] y también en algunas comunidades autónomas. Estas evaluaciones muestran con claridad que ambos conceptos –resultado e impacto– son diferentes. La evaluación de los incentivos a la contratación y al trabajo autónomo es una buena muestra de estas diferencias.

2.2. Un ejemplo de la diferencia entre seguimiento y evaluación de impacto

Tomemos, como ejemplo, la tarifa plana para trabajadores autónomos. De acuerdo con la evaluación de la AIReF, entre 2013 y 2018, se dedicaron 3.500 millones

[7] Véase Malo y Cueto (2015) y Mato (2010).

de euros a reducciones y bonificaciones a la Seguridad Social en este programa, en torno a 600 millones cada año entre 2014 y 2018. Este gasto llegó a unos 600.000 beneficiarios cada año. ¿Quiere decir esto que hay 600.000 nuevos autónomos cada año debido a que se ha puesto en marcha la tarifa plana? En absoluto. El resultado de la política es que, efectivamente, 600.000 personas se beneficiaron cada año de una reducción o una bonificación en las cotizaciones a la Seguridad Social. En torno al 10% de los autónomos en alta en 2018 eran beneficiarios del programa, un 45% de las altas en dicho año.

Pero, en términos de impacto, es decir, en términos de aumento en la probabilidad de que los desempleados se dieran de alta como autónomos, AIReF (2020) señala que solo se encuentra un efecto positivo en personas desempleadas durante periodos muy cortos y con un empleo temporal previo. No hay ningún efecto en personas desempleadas durante más de un mes, ni en las personas desempleadas con bajo nivel educativo. Es decir, estos colectivos se hubieran dato de alta como autónomos, aunque no se hubiera puesto en marcha la tarifa plana de autoempleo. El programa no ha modificado su comportamiento, no hay relación causal entre haber recibido la tarifa plana y decidir iniciar un negocio. Por tanto, el programa presenta un elevado efecto inercia o peso muerto.

La tarifa plana ha tenido mucho éxito en términos de participación, pero su impacto sobre el empleo autónomo en España es, en el mejor de los casos, pequeño. En términos de análisis coste-beneficio, este resultado implica que la mayor parte de los 600 millones anuales mencionados en el párrafo anterior no han tenido ningún impacto sobre el mercado de trabajo y sería mejor haberlos dedicado a otro tipo de medidas que sí tuvieran efectos.

Los indicadores de los PAPE muestran que evaluar, en sentido amplio, las políticas activas de empleo, no es tarea fácil. La heterogeneidad de las medidas –incluso de aquellas que se enmarcan en un mismo concepto, como puede ser formación– es amplia en términos de colectivo al que van dirigidas, duración, requisitos de participación, entre otras características. Se pueden utilizar múltiples indicadores, cada uno de los cuales tendrá ventajas e inconvenientes. En el caso de las evaluaciones de impacto, se añade la dificultad para construir un grupo de control adecuado. Esta dificultad aumenta si, como es habitual en el caso español, las evaluaciones se plantean *a posteriori* y sin haber estado previstas en la fase de diseño y/o implementación de las políticas. La evaluación de las políticas activas que llevó a cabo la AIReF en 2018 es un buen ejemplo.

Este informe puso de manifiesto la dispersión de los datos y que, a pesar de que existe un catálogo de medidas, bajo una misma denominación pueden coexistir programas distintos. «La falta de información detallada sobre las características y requisitos de los servicios en las comunidades autónomas que se traducen a códigos comunes del SISPE» (AIReF, 2019, p. 75) fue uno de los factores que más dificultó el análisis, ya que hace más compleja la evaluación en, al menos, dos aspectos. Por una parte, si no se conocen las características de los programas a evaluar con detalle ni cuáles son los criterios a partir de los que se decide qué personas participan y

cuáles no, la construcción de un grupo de control adecuado es muy compleja. La calidad del grupo de comparación depende de que podamos establecer con claridad por qué unas personas participaron en las políticas activas y otras no, ya que esto ayuda a explicar los sesgos de selección existentes. Por otra parte, la interpretación de los resultados y las posibles recomendaciones son difíciles de hacer si no se conocen los criterios con los que se ponen en marcha los programas y cómo se desarrollan. En definitiva, para llevar a cabo una evaluación de impacto es necesaria una «descripción detallada del contenido de cada una de las actuaciones» a analizar (Felgueroso *et al*., 2018, p. 187).

Los objetivos que se plantearon inicialmente en la evaluación de la AIReF eran muy ambiciosos, puesto que evaluar medidas que se ejecutan a nivel autonómico requiere disponer de un sistema de información con una coordinación muy eficiente. La realidad mostró que los datos disponibles no permitían llevar a cabo esta evaluación, como se verá más adelante.

Sin embargo, aunque no es posible disponer de esa información homogénea a nivel nacional, sí que hay comunidades autónomas en las que se están llevando a cabo evaluaciones con grupo de control. Casi todas ellas, en mayor o menor medida y con mayor o menor difusión, han llevado a cabo evaluaciones de impacto puntuales o las han institucionalizado en planes de evaluación, muchas de las cuales pueden consultarse en las webs de los servicios públicos de empleo autonómicos. Utilizar esas evaluaciones es un elemento imprescindible en el desarrollo de la evaluación en nuestro país.

3. La evaluación en el marco del Fondo Social Europeo

Uno de los elementos que ha impulsado la evaluación de las políticas de empleo en España durante los últimos años ha sido la percepción de fondos europeos por parte de las Administraciones públicas. La cofinanciación de la UE se ha canalizado a través de programas operativos del Fondo Social Europeo, de ámbito tanto regional como plurirregional. Durante el periodo de financiación 2014-2020, los grandes programas plurirregionales han sido los de Empleo, Formación y Educación; Inclusión Social y Economía Social; y Empleo Juvenil. A ellos se les unen 19 programas operativos regionales, correspondientes a las 17 comunidades autónomas y a Ceuta y Melilla.

Las políticas europeas traen consigo protocolos de ejecución y de evaluación. El Reglamento 1303/2013 de la Unión Europea establece la obligación de presentar a la Comisión Europea informes anuales de ejecución de los programas operativos, que son aprobados previamente por los respectivos comités de seguimiento. Respecto a la evaluación, cada programa tiene su plan de evaluación específico, que se enmarca en el Plan Estratégico Nacional de Evaluación del FSE 2014-2020, y que incorpora las directrices establecidas por la Unión Europea.

Entre las exigencias europeas para la evaluación de sus programas figura que las entidades evaluadoras sean externas e independientes, lo que ha dado lugar a contrataciones con consultoras y, en contadas ocasiones, a colaboraciones contractuales con universidades. Las evaluaciones se realizan de forma periódica, aunque no todos los años. La mecánica de las evaluaciones del FSE consiste en que cada Administración gestiona la contratación de las actividades de evaluación de los programas bajo su competencia, ofreciendo los datos necesarios y exigiendo la calidad que considere oportuna, dentro de los márgenes establecidos por el plan de evaluación, por la Unidad Administradora del FSE en España y por la normativa europea.

Por tanto, el modelo de evaluación que impulsan los programas europeos es claramente descentralizado, lo que da lugar a una variedad y diversidad de evaluaciones. La existencia de los planes de evaluación garantiza que haya una mínima comparabilidad entre las evaluaciones de los diferentes programas. El carácter descentralizado facilita que la calidad de los datos utilizados en las evaluaciones pueda ser elevada, y que los problemas de gestión y de interpretación de la información, que son normales en toda evaluación, puedan ser canalizados adecuadamente entre la administración gestora y la entidad evaluadora.

Las directrices europeas respecto a la evaluación de actividades del FSE, además de establecer unos mínimos que permitan la ya mencionada comparabilidad, incluyen otras orientaciones que, sin ser de obligado cumplimiento, marcan pautas deseables para la evaluación. Así, durante los últimos periodos de financiación, la Comisión Europea ha ido incluyendo progresivamente orientaciones para que la parte de las evaluaciones dedicada al impacto incorpore enfoques contrafactuales (Comisión Europea, 2013; 2015). En este sentido, el enfoque descentralizado de la evaluación facilita que cada Administración decida la conveniencia de incluir este enfoque para alguna de las medidas realizadas, en función del tipo de medida, de la disponibilidad de datos, del horizonte temporal de la evaluación y de sus propios objetivos, entre los cuales tampoco cabe excluir la oportunidad política.

Por consiguiente, el hecho de que la Unión Europea establezca los contenidos mínimos, pero no los máximos, para la evaluación de las políticas de empleo que contribuye a financiar, significa una de las grandes ventajas del modelo de evaluación descentralizada: dar cabida a la iniciativa de las Administraciones gestoras en materia de evaluación. Además, como queda dicho, la Comisión valora las evaluaciones de impacto con enfoque contrafactual[8].

Este planteamiento contribuye a explicar la aparición, durante los últimos años, de un número creciente de evaluaciones de tipo contrafactual referidas a programas europeos en España. Por ejemplo, a escala estatal, en 2018 se llevó a cabo la segunda evaluación de la Iniciativa de Empleo Juvenil, que incluyó por vez primera un análisis del impacto de las medidas y un enfoque contrafactual (Ministerio de Trabajo, Migraciones y Seguridad Social, 2018).

[8] Buena prueba del interés de la Comisión Europea en intensificar las evaluaciones de impacto es el Informe publicado en julio de 2019, que analiza la posibilidad de realizar este tipo de estudios en distintos países, entre ellos, España (Comisión Europea, 2019).

Independientemente de la evaluación que el FSE requiere, por lo que respecta a las comunidades autónomas, la proliferación de evaluaciones con contrafactual es notable, aunque no alcanza a todas ellas, como se observa en el Cuadro 1. Dos comunidades autónomas (Cataluña y Navarra) tienen planes propios de evaluación de las políticas activas, en los que se contempla expresamente la evaluación con grupo de control. Además, esta proliferación coincide con el hecho de que, en comunidades como Cataluña, la evaluación de impacto de las políticas de empleo es especialmente intensa. También en Madrid, anualmente, se evalúa el impacto de las políticas activas. En otros casos, las iniciativas no tienen una periodicidad establecida, pero sí se ha realizado alguna evaluación de impacto. A su vez, esta intensidad parece independiente de la existencia de planes autonómicos de evaluación diferentes a los ya citados, que vienen impuestos por el FSE (Cuadro 1).

En definitiva, la evaluación de políticas de empleo en el marco del FSE, al ser descentralizada, remite al hecho de que algunas Administraciones van más allá de los mínimos establecidos por la UE, desarrollando sus propias evaluaciones de impacto, con o sin plan regional de evaluación. A su vez, la evaluación descentralizada de programas del FSE bajo un paraguas común, pero flexible, establece el ámbito de la evaluación al mismo nivel del ámbito de la gestión, mostrando, por tanto, un elemento de contraste muy claro con el enfoque del PAPE que se presentó en la sección anterior. Finalmente, este marco facilita la iniciativa de los gestores, lo que a su vez podría favorecer la difusión de buenas prácticas de evaluación.

4. Datos y fuentes de información para la evaluación

No hay duda de que todos los estamentos implicados en la elaboración y ejecución de las PAE coinciden en considerar que la evaluación de estas políticas es imprescindible para determinar cuál es su impacto y el grado de cumplimiento de los objetivos que persiguen, siendo el mejor camino para medir su eficiencia y eficacia.

En España, la práctica de la evaluación no se ha institucionalizado globalmente, de manera que, aunque recientemente se han llevado a cabo diversos estudios de investigación, sobre todo de carácter académico y con técnicas cuasiexperimentales, solo se dispone de conclusiones parciales y dispersas, que no ayudan a la toma de decisiones.

El modelo ideal para poder realizar una evaluación es aquel que permite identificar qué parte de los resultados obtenidos se deben única y exclusivamente al hecho de haber recibido algún «tratamiento» por parte de una política activa, es decir, que permite establecer una relación de causalidad entre el programa y sus resultados[9].

[9] Un magnífico ejemplo de la relevancia de los análisis basados en las relaciones de causalidad entre sucesos es la reciente concesión del premio Nobel de Economía a David Card, Joshua Angrist y Guido Imbens por su contribución al desarrollo de instrumentos analíticos para identificar qué programas pueden considerarse causantes, y en qué grado, de los resultados observados.

CUADRO 1

PANORAMA DE LAS EVALUACIONES DE POLÍTICAS ACTIVAS DE EMPLEO EN LAS CC. AA.

Comunidad autónoma	Plan propio de evaluación*	Alguna evaluación de impacto (2017-2021)	Alguna evaluación de impacto (2017-2021) con enfoque contrafactual
Andalucía	No	Sí	Evaluación específica sobre la Iniciativa de Cooperación Social y Comunitaria [Programa Empleo 30+]
Aragón	No		
Principado de Asturias	No	Sí	Informe de evaluación. PO FSE Asturias 2014-2020 (2019) Informe de evaluación. PO FSE Asturias 2014-2020 (2017)
Baleares	No	Sí	Evaluación de los objetivos/resultados del PO FSE de Balears (2019) Evaluación de impacto del Programa VISIBLES dirigido a desempleados maduros (2018) Evaluación de políticas activas de ocupación: los programas MIXTS de formación y ocupación (2018)
Comunidad Valenciana	No	Sí	Informe de evaluación de los objetivos/resultados del Programa Operativo FSE 2014-2020 de la Comunidad Valenciana (2019)
Canarias	No	Sí	Evaluación de los objetivos/resultados del PO FSE 2014-2020 de Canarias (2019)
Cantabria	No	No	
Castilla-La Mancha	No	No	
Castilla y León	No	No	

NOTA: * Excluyendo los planes de evaluación específicos de los Programas Operativos del FSE.
FUENTE: Elaboración propia.

CUADRO 1 *(Continuación)*

PANORAMA DE LAS EVALUACIONES DE POLÍTICAS ACTIVAS DE EMPLEO EN LAS CC. AA.

Comunidad autónoma	Plan propio de evaluación*	Alguna evaluación de impacto (2017-2021)	Alguna evaluación de impacto (2017-2021) con enfoque contrafactual
Cataluña	Sí	Sí	Evaluación de impacto del programa Nuevas Oportunidades de Empleo 2015-2017 (2020) Evaluación del programa de formación en Áreas prioritarias (FOAP), ejecución 2016-2019 (2020) Evaluación de diseño, implementación e impacto del programa de formación de los CIFO, ejecución 2016-2019 (2020) Informe comparativo de los programas de formación destinados a personas en situación de paro, ejecución 2016-2019 (2020) Evaluación de impacto del programa Trabajo y Formación – Línea parados de larga duración. Convocatoria 2016 (2020) Evaluación de impacto del programa Trabajo y Formación (convocatoria 2015) en el marco del PO FSE Cataluña 2014-2020 (2017) Evaluación de implementación e impacto del programa TLN Mobilicat, convocatoria 2015 (2018)
Extremadura	No	Sí	Evaluación de los objetivos/resultados del PO FSE 2014-2020 de la comunidad autónoma de Extremadura (2019)
Galicia	No	Sí	Evaluación de los objetivos/resultados de las prioridades del PO FSE Galicia 2014-2020 (2019)
La Rioja	No	No	
Madrid	No	Sí	Evaluación del Impacto Individual de Políticas Activas de Empleo 2018 (2019)
Murcia	No	Sí	Informe de evaluación de los objetivos/resultados del Programa Operativo FSE 2014-2020 Región de Murcia (2019)
Navarra	Sí	No	
País Vasco	No	Sí	Informe de evaluación de los objetivos/resultados del Programa Operativo FSE 2014- 2020 del País Vasco (2019)

NOTA: * Excluyendo los planes de evaluación específicos de los Programas Operativos del FSE.
FUENTE: Elaboración propia.

Este tipo de procedimiento no es sencillo, pero cabe la posibilidad de que se diseñe una determinada política de manera que se pueda poner en marcha un experimento previo. Por ejemplo, mediante experiencias piloto que permitan, de manera previa al lanzamiento de un programa, medir la eficacia de dichas medidas. En este sentido, los experimentos en cuestiones relacionadas con la pobreza y el desarrollo, de Duflo y Banerjee constituyen una referencia (Banerjee & Duflo, 2009; Duflo, 2017), si bien estas aplicaciones estuvieron precedidas de otras muchas en el ámbito de la economía laboral y, específicamente, en el de las políticas activas de mercado de trabajo (véase Heckman, Lalonde & Smith, 1999). Desafortunadamente, en Europa, las evaluaciones experimentales en este campo son escasas y se concentran principalmente en los países nórdicos[10].

La dificultad de realizar este tipo de evaluaciones, que debe ser prevista en el momento de diseño de la política, puede ser solventada en ocasiones gracias a la existencia de metodologías estadísticas conocidas como cuasiexperimentales, que utilizan los datos disponibles, como son los procedentes de registros administrativos, para extraer información, que permiten interpretar los resultados como si fuesen fruto de la comparación de los beneficiarios de una política con un grupo de control definido mediante un experimento. No obstante, poder desarrollar estas metodologías adecuadamente requiere de un conocimiento detallado del significado y recolección de los datos para que los resultados del análisis no estén invalidados debido a los sesgos introducidos por otras variables que escapan al control del investigador.

En España, la ausencia de evidencia experimental es una grave limitación, porque la realización de experimentos permite conocer el impacto de las políticas activas contando con información de calidad recogida explícitamente para ello. Las evaluaciones cuasiexperimentales están concentradas en muy pocas comunidades autónomas, siendo conveniente que se extiendan a todos los niveles, lo que podría ser impulsado desde el SEPE, quien además podría coordinarlas a través de la organización administrativa existente de los SPE. El coste sería reducido, pues los datos ya existen y lo único que se necesita es una extracción correcta de las bases de datos disponibles y un análisis que, hoy en día, es mucho más fácil de acometer gracias a diferentes paquetes estadísticos (Malo & Cueto, 2015). Los investigadores d'Hombres y Santangelo (2019) presentan las ventajas del uso de los datos administrativos para hacer evaluación de impacto de manera sistemática, porque proporcionan gran cantidad de datos y de información con un coste de utilización muy bajo, en comparación con el coste derivado de la realización de encuestas. Destaca, sin embargo, que España no aparece en el grupo de países que tienen publicados análisis de evaluación utilizando modelos contrafactuales.

Ante esta evidente escasez de procesos de evaluación adecuados y completos, no podemos dejar de llamar la atención del que podría ser el principal motivo que lo explica.

[10] Véase, por ejemplo, el metaanálisis realizado por Card, Kluve y Weber (2018).

La posibilidad de poner en práctica estos procesos de evaluación en España está condicionada por la disponibilidad de datos adecuados para tal fin. No se puede afirmar que no haya información, sino que es complicado acceder a ella y, en ocasiones, es incompleta. Además, incluso cuando existe y está disponible, resulta difícil manejarla para analizar lo que se está buscando.

Partiendo del supuesto de que las PAE dejan claramente definidos, estructurados y concretados los objetivos que persiguen, las fuentes de datos utilizadas en la evaluación han de contener información suficiente para establecer cuál es la variable que permite medir el resultado, así como para aislar qué efectos se han debido exclusivamente al tratamiento recibido por la PAE, de aquellos que se alcanzan con independencia de ella. Como ya se ha mencionado anteriormente, resulta imprescindible llegar a identificar a la población que recibe el tratamiento de la PAE y diseñar un grupo de control que tenga idénticas características que el grupo beneficiario, pero que no ha recibido ningún tipo de tratamiento.

Por otra parte, una vez que estén bien definidos y validados los elementos y cuestiones que han de ser analizados, hay que identificar los criterios para la estandarización de conceptos y los indicadores que permitan hacer el seguimiento y la posterior evaluación. En este proceso, es importante identificar a los interlocutores y su intervención en la oferta y puesta en marcha de los servicios y programas. El conocimiento referido al detalle de la implementación y el diseño de la evaluación requieren de una estrategia que ayude a recopilar toda la información necesaria para su realización. No solo es necesario definir la idoneidad de la estrategia propuesta en relación con las necesidades del mercado de trabajo, sino el diseño del procedimiento que se va a utilizar en la posterior evaluación.

A modo de ejemplo, se puede considerar que una variable resultado puede ser en qué medida mejora la empleabilidad o el desempeño laboral de un individuo tras ser tratado por una PAE. Ahora bien, medir el grado de empleabilidad o el desempeño laboral no es tarea fácil[11], hay que poder determinar qué se entiende y cómo se mide, porque no solo depende de que se logre salir del desempleo, sino del tiempo que se tarda en lograrlo, de la proporción de tiempo que el individuo está empleado, el nivel de ganancias acumuladas, el tiempo transcurrido hasta el primer empleo, del tipo de contrato con el que se accede, de la duración de dicho contrato, del nivel de cualificación y su adecuación con el empleo, entre otras.

4.1. El encaje temporal de la evaluación

Si la evaluación está prevista desde la fase de diseño de la política, debe decidirse la metodología a emplear y, con ello, también el protocolo de recogida de información en la que se basará la evaluación. Como ya se ha mencionado, una vez

[11] Toharia *et al*. (2006) analizan con detalle de qué manera hay que tratar los datos registrados en las demandas de empleo en el SEPE para poder establecer alguna medida de la empleabilidad.

que estén bien definidos y validados los elementos y las cuestiones que han de ser analizadas, hay que identificar los criterios para la estandarización de conceptos y los indicadores que permitan hacer el seguimiento y la posterior evaluación. En este proceso es importante identificar a los interlocutores y su intervención en la oferta y puesta en marcha de los servicios y programas. El conocimiento referido al detalle de la implementación y el diseño de la evaluación requieren de una estrategia que ayude a recopilar toda la información necesaria para su realización. No solo es necesario definir la idoneidad de la estrategia propuesta en relación con las necesidades del mercado de trabajo, sino el diseño del procedimiento que se va a utilizar en la posterior evaluación. En este sentido, el papel de los agentes participantes, de las oficinas de empleo y, por ejemplo, de los centros de formación resulta muy relevante para poder tener el control de las medidas de eficiencia y eficacia de los servicios y programas prestados.

Lamentablemente, la decisión sobre evaluar PAE se toma, en muchas ocasiones, cuando la política ya se ha implementado y, así, gran parte de las dificultades para evaluar reside en que la recogida de datos no ha sido adecuada a la necesidad del programa a evaluar. Estas dificultades se pusieron claramente de manifiesto en la primera evaluación que intentó hacer la AIReF, donde se indica que «los datos suministrados han sido insuficientes y de una calidad mejorable. La evaluación completa de la eficacia de las políticas activas por comunidad autónoma no ha sido posible» (AIReF, 2019, p. 6).

Los objetivos que se planteaban inicialmente eran muy ambiciosos. Sin embargo, finalmente, se evaluaron únicamente dos programas: medidas de diagnóstico individualizado en Aragón y el PREPARA, cuya información era recogida por el SEPE a nivel nacional. Emplearemos el plan PREPARA para mostrar la dificultad de configurar un grupo de control adecuado, en la línea de lo planteado anteriormente.

El plan PREPARA entró en vigor en febrero de 2011[12], con el fin de apoyar la cualificación profesional de las personas que agotaban su protección por desempleo, mediante acciones de políticas activas de empleo y la percepción de una ayuda económica de apoyo. Los requisitos para recibir la ayuda variaron a lo largo del tiempo[13], lo que ya es una dificultad en sí misma, dado que los grupos de tratamiento y de control deben tener en cuenta los cambios realizados. Por otra parte, se disponía de muy poca información sobre los servicios ofrecidos a los beneficiarios. En el informe de la AIReF se indica que no se incluye en el análisis la duración de dichos servicios por la mala calidad del dato. Dada la naturaleza del programa y su apuesta por combinar la ayuda con acciones de políticas activas, estas últimas eran un aspecto clave en la evaluación. Es más, la comparación entre grupos que recibieran

[12] Real Decreto-ley 1/2011, de 11 de febrero, de medidas urgentes para promover la transición al empleo estable y la recualificación profesional de las personas desempleadas.

[13] Real Decreto-ley 14/2017, de 6 de octubre, por el que se aprueba la reactivación extraordinaria y por tiempo limitado del programa de recualificación profesional de las personas que agoten su protección por desempleo.

distintas medidas (orientación, formación, u otras) podría haber aportado resultados muy interesantes sobre qué tipo de programas podrían ser más útiles para la inserción laboral de los beneficiarios.

En el análisis del PREPARA se incide en que genera desincentivos a incorporarse al mercado laboral en los seis meses que dura la prestación. Aquí, podemos señalar la importancia de los aspectos subjetivos de valoración de resultados. Dado que el colectivo destinatario está formado por personas desempleadas de larga duración, principalmente con baja cualificación, ¿podemos esperar resultados positivos a corto plazo en una coyuntura económica de crisis? Por otra parte, imaginemos que las personas están participando en actividades formativas y que, dado que reciben una ayuda, prefieren seguir formándose a aceptar un empleo que, probablemente, sea de muy corta duración. ¿Este desincentivo lo calificamos como «negativo»? Por supuesto, lo que estamos planteando son hipótesis que, con datos de calidad, podrían contrastarse.

Por último, hay que señalar las diferencias que se encontraron entre comunidades autónomas que aportan información para el aprendizaje mutuo de gran interés. Los resultados trascurridos dos años mostraban que, en tres casos, el impacto era negativo, en dos casos positivo y, en el resto, nulo. Además de vincular esta información con la diferente situación de los mercados de trabajo regionales, la explicación también podría estar vinculada con la distinta supervisión de la búsqueda activa de empleo o las distintas medidas en las que se participaba. Desafortunadamente, la información para seguir ahondando en los resultados del programa no estaba disponible.

Respecto a las medidas de diagnóstico individualizado en Aragón, son un buen ejemplo de qué se puede hacer si se cuenta con información adecuada. El informe de la AIReF disponía de información sobre qué personas habían recibido servicios tras el diagnóstico individualizado. Los datos permitían caracterizar a las personas beneficiarias en términos de trayectoria laboral y acciones del servicio público de empleo en las que habían participado. En definitiva, la calidad de la información sobre ese programa permitió llevar a cabo la evaluación de impacto.

4.2. El potencial de los registros administrativos

Los datos administrativos procedentes de los registros de la Seguridad Social o del Servicio Público de Empleo son una importante fuente de información con un coste relativamente bajo que, además de contener una cantidad elevada de información, permiten llegar a unidades de análisis territoriales pequeñas. Su relevancia es tal que estos datos administrativos podrían ser elevados a la consideración de «bien público». Si fuese fácil y sencilla su disponibilidad para las personas residentes en un país, habría un mayor y mejor conocimiento de la sociedad, además de proporcionar la tan demandada e imprescindible transparencia en favor de la democracia.

Estas bases de datos administrativas han sido utilizadas previamente por investigadores que han evaluado políticas de empleo, principalmente incentivos a la

contratación, y que han puesto de manifiesto sus fortalezas y debilidades como fuente de información[14].

Desde 2005, España dispone del Sistema de Información de los Servicios Públicos de Empleo (SISPE), en el cual se recoge importante información sobre los demandantes de empleo que están inscritos en las oficinas públicas de empleo. Asimismo, se incorpora información sobre los Servicios y Programas comunes prestados por el Estado y propios de las comunidades autónomas. Y también integra el historial de servicios recibidos por las personas inscritas en algún momento y se dispone de información referida a determinados programas, como el PREPARA. Por otro lado, el sistema incluye toda la información recogida en el registro de contratos, o en los sistemas que inscriben a los perceptores de prestaciones por desempleo.

La evaluación de la eficacia de las políticas activas de empleo parte de la selección de las personas inscritas en el SEPE que son demandantes de empleo que realmente quieren trabajar, porque hay demandantes inscritos por otras razones relacionadas con los posibles derechos que el registro puede reportarles tanto en el corto, como en el medio y largo plazo. De ahí que dentro del SISPE se clasifique a los colectivos inscritos en función de que realmente estén buscando un puesto de trabajo porque quieren trabajar. El problema es que, seguramente, puede haber personas clasificadas como paradas que en realidad no se encuentren disponibles para ocupar un empleo. Además, la disponibilidad para trabajar es una cuestión que no puede ser observada, que depende de las personas, sus actitudes y sus condiciones, lo que sitúa a la disponibilidad en un plano subjetivo de muy difícil medición. De ahí que haya ciertos problemas con este tipo de datos, que sin lugar a dudas es inmensamente rico e interesante, que levantan preocupaciones y obligan a ser cautos tanto en su uso como explotación e interpretación.

Además, hay que añadir las numerosas reticencias y temores por parte de los agentes gestores de los registros ante la necesidad de guardar la confidencialidad que, en ocasiones, se debe más a un exceso de celo que a un problema real, pues podrían resolverse, y de hecho así se hace en otras fuentes, con la debida anonimización de las bases de datos y la agregación de determinadas categorías.

Por otra parte, el proceso de solicitud es complicado, y se complica más cuando resulta imprescindible la unión de diversas bases de datos administrativos, puesto que hay que implicar a diversos gestores que, incluso perteneciendo a la misma institución, utilizan usos y costumbres dispares.

Otra magnífica fuente de información que pone a disposición de los investigadores datos de la Tesorería General de la Seguridad Social es la Muestra Continua de Vidas Laborales (MCVL). Esta base de datos es resultado de una operación estadística emprendida por la Dirección General de Ordenación de la Seguridad Social, con el objetivo de explotar la enorme riqueza de información que existe en los registros

[14] Véase, por ejemplo, Cebrián *et al*. (2011), trabajo en el que colabora Luis Toharia, investigador con una participación decisiva para la puesta en marcha de la renovación del sistema de registros del Servicio Público de Empleo, así como pionero en enlazar diferentes fuentes de registros administrativos.

informatizados de la Seguridad Social, no solo desde el punto de vista de la afiliación durante los periodos de empleo, sino también en lo referente a las prestaciones sociales recibidas[15].

El primer año de su publicación fue 2004, pero contiene el número de episodios de afiliación iniciados en cada año, ya sea por empleo por cuenta ajena o propia, por una prestación por desempleo o por alguna pensión. Además, proporciona todo el historial de la vida laboral que haya registrado para cada una de las personas que forman parte de la muestra, lo que supone una de sus principales ventajas al permitir obtener datos longitudinales para el seguimiento de la trayectoria laboral de las personas de la muestra y así tener un panel con los historiales de todos los participantes. Ahora bien, recientemente se ha incorporado a la solicitud de dicha base el requisito de que los datos sean destruidos anualmente, lo que, si no se modifica, hará inviable su uso para el análisis longitudinal.

En algunas ocasiones, ha sido necesario solicitar el enlace de diferentes fuentes administrativas, como las de los registros del SEPE y los registros de la Seguridad Social de la MCVL[16]. En todas ellas el procedimiento ha sido largo y tedioso y, a veces, casi imposible, aunque también es cierto que ha ido mejorando con la práctica y experiencia, hasta tal punto que, en ocasiones, se oyen voces comentando el interés de las partes en que el enlace se convierta en algo habitual. Esto evitaría que una vez finalizado el estudio, los datos hayan de ser destruidos impidiendo que se vuelva a utilizar dicho enlace.

Un buen ejemplo de enlace de datos administrativos lo podemos encontrar en la unión del fichero de contratos registrados en el SEPE y la MCVL. El resultado que se obtiene es uno de los muchos ejemplos que confirman lo conveniente que es la colaboración entre Administraciones para poder llevar a cabo cualquier análisis de evaluación. La principal ventaja de este enlace es que permite complementar la información de ambas bases de información. Por un lado, el fichero de contratos registrados no permite identificar la duración efectiva del contrato, solo el momento en el que se inicia y, en todo caso, se podría conocer si el individuo tuvo con anterioridad o posterioridad otro contrato, de modo que se podría calcular el total de contratos firmados a lo largo de un periodo, pero en ningún caso se podrá saber el tiempo transcurrido en cada tipo de contrato ni entre la finalización de uno y el comienzo del otro. En cambio, esta información es posible conocerla con los datos aportados por la MCVL, ya que figura el momento del inicio de la relación laboral, el momento de finalización, si es que lo hay, e incluso se puede identificar si el individuo ha cambiado de contrato dentro de la misma relación laboral, pudiendo deducirse el tiempo transcurrido desde el inicio de dicha relación y el tiempo pasado en los diferentes contratos que se hayan ido encadenando a lo largo de ella, e incluso el tiempo fuera del sistema

[15] No se encuentran incluidos los demandantes de empleo cuando no reciben prestaciones y los inactivos (distintos de los pensionistas), además de los trabajadores que tienen un sistema de previsión social distinto de la Seguridad Social (los funcionarios de Clases Pasivas) o no tienen ninguno (como quienes trabajan en la economía informal).

[16] Cebrián, Moreno y Toharia (2011a); Cebrián, Moreno y Toharia (2011b); Cebrián y Moreno (2012).

de afiliación. Sin embargo, la MCVL, a pesar de incorporar datos procedentes del padrón aportados por el INE, proporciona una información sobre el nivel de estudios que no tiene por qué corresponder con la del individuo a la firma de cada contrato que haya podido tener a lo largo de su vida laboral. En cambio, dicha información sí se encuentra disponible en el registro de contratos. Lo mismo ocurre con el tamaño de la plantilla, ya que dicha información se registra en la MCVL según consta en la Seguridad Social en el momento en el que se extrae la muestra de referencia, no cuando se registra el contrato. Además, el fichero de contratos contiene información detallada sobre el tipo de ocupación a la que se refiere el contrato, en cambio, este detalle no se recoge en la Seguridad Social. Por ello, gracias al enlace se combina toda la información de la Seguridad Social con la aportada por el registro de contratos. Por otro lado, el registro de contratos aporta mucho detalle sobre la normativa que regula la contratación, si el contrato está sujeto a algún tipo de bonificación o deducción financiada por el Servicio Público de Empleo, así como el tipo de indemnización por despido improcedente que, antes de la reforma laboral de 2012, podía ser de 33 días por año trabajado con un máximo de 24 mensualidades o de 45 días por año trabajado con un máximo de 42 mensualidades, en función de las características del contrato.

El principal inconveniente es el propio proceso de enlace. Ya que requiere que las Administraciones competentes acuerden cuál será el criterio y procedimiento para realizar la unión y la anonimización del identificador de la persona física y del código de la cuenta de cotización, que son las dos variables básicas para proceder al enlace. Sirva como otro ejemplo de la complejidad para llevar a cabo este enlace que el identificador de la persona física según el criterio que aplica la Seguridad Social y el utilizado por el fichero de contratos son diferentes, de manera que resulta prácticamente imposible identificar adecuadamente a los trabajadores de nacionalidad extranjera por ser diferente el criterio de identificación seguido por la Seguridad Social y por el SEPE. Esta circunstancia impide que se identifiquen adecuadamente aquellos contratos de la población extranjera que forma parte de la MCVL. Por ello, en ocasiones, el estudio se centra exclusivamente en aquellos individuos que en el momento de firmar un contrato poseen la nacionalidad española, con independencia de cuál fuese su país de nacimiento. Evidentemente, esto puede dar lugar a que se elimine casi a un quinto del total de contratos y afiliaciones.

En resumen, desde el punto de vista del análisis y la investigación, resulta absolutamente necesario poder utilizar los datos administrativos con la imprescindible colaboración e implicación directa de sus gestores y propietarios, de manera que se garantice el acceso e interpretación adecuada de los conceptos y registros, elementos base para tener un buen conocimiento y entendimiento de los sistemas de información.

Para evitar retrasos y errores sería muy conveniente que hubiese unidades dedicadas en la planificación y coordinación de la evaluación de las políticas activas en las que estuvieran implicadas y colaboraran las diferentes unidades gestoras de las fuentes, así como las entidades encargadas de realizar las evaluaciones.

5. Conclusiones

En este trabajo se han repasado una serie de aspectos clave en relación con la evaluación del impacto de las políticas activas de empleo en España. A partir de la experiencia de los autores y de un análisis documental de la bibliografía disponible, se han puesto de manifiesto varias cuestiones, comenzando por la constatación de que la normativa en materia de PAE incluye a la evaluación como parte integrante de los Planes Anuales de Política de Empleo, si bien la concreción de los indicadores que abarcan los PAPE resultan insuficientes para captar los verdaderos efectos de las políticas que se llevan a cabo. En este sentido, se han expuesto las carencias de los PAPE en comparación con la necesidad de evaluar el impacto causado por las medidas, disociándolo con claridad de la influencia de otros factores, como el ciclo económico o las diferencias entre diversos mercados de trabajo regionales. Se han puesto de manifiesto, como ejemplo, los significativos efectos de inercia de la tarifa plana para trabajadores autónomos.

A su vez, se ha analizado la evaluación de las políticas activas llevada a cabo por la AIReF en 2019, y los problemas destapados a partir de la misma en relación con las diferencias en las pautas de registro de información por parte de las comunidades autónomas.

Los problemas derivados del marco normativo estatal y de la falta de homogeneidad de la información se han contrastado con el mecanismo de evaluación que utiliza el Fondo Social Europeo, que hace recaer sobre las Administraciones gestoras, ya sean nacionales o regionales, las tareas de evaluación, sometidas a ciertos principios comunes. Se han señalado las ventajas e inconvenientes de este modelo descentralizado, aportando evidencia del incremento creciente del número de evaluaciones de impacto con enfoque contrafactual durante los últimos años.

Adicionalmente, en el artículo se ha abordado el contraste entre la falta de una institucionalización de la evaluación y la existencia de numerosos trabajos académicos que aportan evidencias específicas sobre el impacto de las PAE. No obstante, también cabe señalar que no se ha realizado ninguna evaluación de tipo experimental que, a pesar de tener un coste elevado, presenta ventajas derivadas de poder tener acceso a información ajustada al objetivo de la evaluación mediante encuestas *ad hoc*. Por el contrario, los datos administrativos, aunque su objetivo no es facilitar la evaluación, constituyen una magnífica fuente de información, cuyo coste es más reducido y permiten acceder a unidades de análisis muy desagregadas, incluso a la propia población tratada.

A este respecto, la disponibilidad de datos de calidad resulta crucial, lo que remite a los procedimientos de recogida y tratamiento de la información. Se ha tomado como muestra de estos problemas el hecho de que la evaluación de eficacia de las PAE hecha por la AIReF y citada anteriormente, que ambicionaba un ámbito estatal, hubiese de conformarse en cuanto a análisis de impacto causal con el análisis de un solo programa estatal (PREPARA) y otro de una comunidad autónoma. En contraste, se han presentado las importantes ventajas de la Muestra Continua de Vidas Laborales, de la que también se han señalado algunos inconvenientes específicos.

El conjunto de análisis realizado permite establecer tres comentarios a modo de conclusión, a los que se unen algunas propuestas de mejora de la evaluación de las PAE. Entre los comentarios, primero, cabe subrayar el problema derivado de la falta de colaboración entre las Administraciones que son propietarias de la información y las entidades que podrían llevar a cabo evaluaciones de impacto. A este problema se le suma, en segundo lugar, que muchos datos que podrían publicarse no se publican. En tercer lugar, algo que quizá ayude a entender los dos puntos anteriores, cabe subrayar la falta de homogeneidad de los sistemas de recogida de información de las Administraciones públicas con competencias en la gestión de las PAE.

Entre las propuestas de mejora de la evaluación de políticas de empleo se pueden incluir, sin ánimo de exhaustividad, las siguientes: primero, dados los decepcionantes resultados de las iniciativas de evaluación de PAE a escala estatal, cabría avanzar hacia un modelo descentralizado inspirado en el que utiliza el FSE, donde la responsabilidad de la evaluación recaiga sobre las Administraciones responsables de la gestión de las medidas. Este modelo, a pesar de los problemas de comparabilidad de las evaluaciones, tiene indudables ventajas, visibles ya al aparecer cada vez más evaluaciones de impacto que avanzan conclusiones útiles sobre la causalidad de las medidas de empleo.

En segundo lugar, cabe defender que se avance también hacia un mayor compromiso por parte de las Administraciones públicas que deberían protagonizar o facilitar evaluaciones a escala estatal, como la Seguridad Social, el SEPE y los ministerios con competencias en PAE y en registro de información necesaria para la evaluación.

Por último, una propuesta muy específica que facilitaría la evaluación de las PAE es la posibilidad de vincular las diversas fuentes de datos generados por las Administraciones públicas (al menos Seguridad Social y Servicio Público de Empleo, pero también, Agencia Tributaria, INE), como ya se hace cuando la MCVL se enlaza con datos del INE y de la Agencia Tributaria. Su explotación por parte de las diversas Administraciones y la comunidad científica y académica, gracias a la digitalización y las cada vez más avanzadas tecnologías disponibles, proporcionarán la generación de un mayor y mejor conocimiento sobre el impacto de las PAE.

Por último, cabe mencionar que la aplicación estricta, y a veces muy restrictiva, tanto de la ley de protección de datos[17], como del Reglamento General de Protección de Datos europeo para garantizar la seguridad de los datos de carácter personal que se incluyen en las fuentes de información, limitan notablemente la explotación y uso de estas fuentes por parte de investigadores externos a la Administración pública, lo que impide que se puedan llevar a cabo estudios de evaluación de las políticas públicas o replicar las evaluaciones ya realizadas como garantía de la propia evaluación.

Algunos países europeos con la misma legislación han arbitrado mecanismos de acceso a los datos para facilitar que personal investigador externo a la propia Administración pública pueda proceder a la realización de evaluaciones. Este podría ser

[17] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

el ejemplo de la iniciativa desarrollada por la Seguridad Social, el INE, la Agencia Tributaria y el Banco de España para poner en marcha un sistema de acceso a sus bases de datos con fines científicos. Se han puesto a disposición del personal investigador solamente dos salas de investigación en Madrid y en Barcelona, lo que ayuda a garantizar el secreto estadístico, la confidencialidad de los datos y la protección de los datos personales, pero quizá limite un acceso amplio.

Referencias bibliográficas

AIReF, Autoridad Independiente de Responsabilidad Fiscal. (2019). *Evaluación Del Gasto Público 2018. Estudio Programa Políticas Activas de Empleo*.

AIReF, Autoridad Independiente de Responsabilidad Fiscal. (2020). *Evaluación Del Gasto Público 2019. Estudio Incentivos a La Contratación*.

Banerjee, A. V., & Duflo, E. (2009). The Experimental Approach to Development Economics. *Annual Review of Economics, 1*, 151-178. https://doi.org/10.1146/annurev.economics.050708.143235

Card, D., Kluve, J., & Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. J*ournal of the European Economic Association*, *16*(3), 894-931. https://doi.org/10.1093/jeea/jvx028

Cebrián, I., Moreno, G., & Toharia, L. (2011a). *Evaluación microeconómica de los programas de fomento del empleo indefinido a partir de datos de los servicios públicos de empleo y de la Muestra Continua de Vidas Laborales*. Ministerio de Trabajo e Inmigración. ISBN 9788484173847.

Cebrián, I., Moreno, G., & Toharia, L. (2011b). La estabilidad laboral y los programas de fomento de la contratación indefinida. *Hacienda Pública Española / Revista de Economía Pública, 198*, 103-127. ISSN 0210-1173.

Cebrián, I., & Moreno, G. (2012). La estabilidad de los nuevos contratos indefinidos durante la crisis económica. *Estudios de Economía Aplicada*, *30*(1), 183-208. ISSN 1133-3197.

Comisión Europea. (2013). *Design and commissioning of counterfactual impact evaluations*.

Comisión Europea. (2015). *Programming period 2014-2020. Monitoring and Evaluation of European Cohesion Policy. European Social Fund. Guidance document*.

Comisión Europea. (2019). *Pilot and feasibility study on the sustainability and effectiveness of results for European Social Fund participants using counterfactual impact evaluations. Final Report*. https://op.europa.eu/es/publication-detail/-/publication/84cc9eb9-b33d-11e9-9d01-01aa75ed71a1

d'Hombres B., & Santangelo, G. (2019). Use of Administrative Data for Counterfactual Impact Evaluation of Active Labour Market Policies in Europe: Country and Time Comparisons. In N. Crato, & P. Paruolo (eds.), *Data-Driven Policy Impact Evaluation* (pp. 271-287). Springer. https://doi.org/10.1007/978-3-319-78461-8_17

Duflo, E. (2017). The Economist as Plumber. *American Economic Review*, *107*(5), 1-26. https://doi.org/10.1257/aer.p20171153

Felgueroso, F., García-Pérez, J. I., & Jiménez-Martín, S. (2018). *Perfilado Estadístico: Un Método Para Diseñar Políticas Activas de Empleo*. Fundación Ramón Areces. FEDEA.

Heckman, J. J., Lalonde, R., & Smith, J. (1999). Chapter 31. The economics and econometrics of Active Labour Market Policies. *Handbook of Labor Economics, 3*(A), 1865-2097.

Malo, M. A., & Cueto, B. (2015). El impacto de las políticas activas de mercado de trabajo en España. *Documentación social*, 178, 105-120.

Mato, F. J. (2010). La formación en España desde una perspectiva comparada. Balance y propuestas de mejora. *Papeles de Economía Español*a, 124, 266-280.

Ministerio de Trabajo, Migraciones y Seguridad Social. (2018). *II Evaluación de la Iniciativa de Empleo Juvenil. Informe de Evaluación*. Subdirección General de Programación y Evaluación del Fondo Social Europeo.

Toharia, L., Pérez Infante, J. I., & Prudencio, C. A. (2006). *La ocupabilidad de los parados registrados y la tasa de cobertura de las prestaciones por desempleo. Informe para el Servicio Público de Empleo Estatal*.

Vila, J. E., Gómez, Y., Martínez, J., & Ferri, J. L. C. (2020). Evaluación de políticas públicas con métodos cuasi-experimentales: una aplicación a las políticas activas de empleo en la Comunidad Valenciana. *Cuadernos económicos de ICE*, 99, 11-24.

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Evaluación de impacto de políticas activas de empleo para colectivos de difícil inserción laboral*

Active labor market policies for workers with difficult job placement experiences: an impact evaluation

Yolanda Fátima Rebollo-Sanz
Universidad Pablo de Olavide

José Ignacio García Pérez
Universidad Pablo de Olavide y FEDEA

Resumen

En este trabajo se realiza un ejercicio de evaluación de impacto de una política activa típicamente dirigida a parados de difícil inserción laboral. Se trata de las denominadas iniciativas locales de empleo y, concretamente, de dos medidas desarrolladas en Andalucía durante los años 2016-2018: los programas Emple@Joven y Emple@30+. Utilizando métodos de emparejamiento se obtiene que el programa Emple@Joven, dirigido a parados menores de 30 años, no aumenta los niveles de empleabilidad para los beneficiarios del programa y que el programa Emple@30+, dirigido a parados mayores de 30 años, solo favorece la empleabilidad a los parados sin experiencia laboral o aquellos que han estado fuera del mercado de trabajo durante más de 24 meses. Por tanto, estos programas, que ofrecen una experiencia laboral de entre 3 y 6 meses, solo parecen ser efectivos para colectivos muy determinados con especial dificultad para acceder por sí mismos al mercado de trabajo. Este ejercicio de evaluación apunta la necesidad de perfilar mejor a los beneficiarios de las políticas activas de empleo y ofrecer a estos solo las que se estime puedan ofrecer un mayor impacto sobre su nivel de empleabilidad, ayudándoles a una mejor y más estable inserción en el mercado de trabajo.

Palabras clave: *evaluación de políticas activas de empleo, iniciativas locales de empleo, propensity score matching.*

Clasificación JEL: *J24, J68, H53.*

Abstract

In this paper, an impact assessment exercise of an active labour market policy aimed at hard-to-place unemployed is conducted. These are the so-called local employment initiatives and in particular, two measures developed in Andalusia during the years 2016-2018: the Emple@Joven and Emple@30+ programs. Using matching methods, it is found that the Emple@Joven program, aimed at unemployed people under the age of 30, did not increase the employment levels for the beneficiaries of the program and that the Emple@30+ program, aimed at unemployed people over the age of 30, only significantly helped the unemployed people without working experience or those who had been out of the labour market for more than 24 months. Therefore, apparently, these measures that provided a working experience between 3 and 6 months, only seemed to be effective for very specific groups with special difficulty in accessing the labour market on their own. These results, thus, point to the need of a better ex-ante profiling of active employment policies' recipients, offering them only to those who are expected to gain the most in terms of their level of employability, offering them a better and more stable insertion in the labour market.

Keywords: *active employment policy evaluation, local employment initiatives, propensity score matching.*

* Los autores agradecen a la Consejería de Empleo, Formación y Trabajo Autónomo de la Junta de Andalucía la cesión de los datos necesarios para poder realizar este trabajo. También agradecen la financiación recibida de la Agencia Estatal de Investigación (Ministerio de Ciencia e Innovación), proyecto PID2019-104452RB-I00 y del Fondo Europeo de Desarrollo Regional y la Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía, proyectos P18- RT-2135 y CV20-35470. Cualquier error que pudiera incluir el texto es de la exclusiva responsabilidad de sus autores.

DOI: https://doi.org/10.32796/cice.2021.102.7315
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

Es común encontrar informes y documentos de política económica que califican las políticas activas de empleo (PAE) como una herramienta fundamental en la lucha contra el desempleo. De hecho, el papel de estas políticas ha sido y es crucial en Europa como estrategia para mejorar la empleabilidad. Desde que se aprobó por primera vez la Estrategia Europea para el Empleo (EEE) en 1997, han aumentado considerablemente los presupuestos para los programas dirigidos al fomento del empleo de colectivos específicos (jóvenes, parados de larga duración, mujeres, personas en riesgo de exclusión social, etc.). Sin embargo, en pocas ocasiones dichas recomendaciones están sustentadas por un conocimiento claro del impacto de cada una de estas políticas sobre el acceso al empleo o los ingresos futuros de los beneficiarios potenciales. Si bien es cierto que a nivel internacional se cuenta cada vez con más evaluaciones del impacto de las PAE (hay estudios que sistematizan los resultados con un metaanálisis, Card, Kluve & Weber, 2010; Kluve, 2010; o Card *et al*., 2018), en España se sigue padeciendo de una escasez de evaluaciones de las distintas PAE en general (algunas excepciones son García-Pérez & Rebollo-Sanz, 2009a, 2009b; Cueto & Mato, 2009; Arranz *et al*., 2013; De la Rica, 2015). Y aún más extrema es la escasez de ejercicios de evaluación de impacto de las iniciativas locales de empleo, objeto del presente artículo.

En este sentido, este trabajo presenta los resultados de un ejercicio de evaluación de impacto de las iniciativas locales de empleo que se ejecutaron en la Comunidad Autónoma de Andalucía durante los años 2016-2018. Esta política se desarrolló bajo el nombre de Proyectos de Cooperación Social y Comunitaria y tenía como objetivo ofrecer experiencias de empleo temporales a sus beneficiarios, prioritariamente individuos parados con dificultades de inserción laboral, de cara a mejorar su empleabilidad futura y su completa inserción en el mercado de trabajo.

Este tipo de actuaciones han sido muy usadas en España durante los últimos 10-15 años, pero fue a raíz de la última gran crisis económica cuando surgieron con especial énfasis este tipo de medidas en distintas comunidades autónomas. En efecto, en el año 2008 comenzó lo que vino a llamarse la Gran Depresión. Cinco años después, la tasa de paro era todavía muy elevada, con niveles que superaban, en el año 2015, el 20 % en el conjunto de España y el 29 % en Andalucía. La tasa de paro juvenil era mucho más abultada, con niveles que superaban el 45 % en España y el 55 % en Andalucía. Pero la manifestación más dramática de la dureza de esta crisis fue el desempleo de larga duración. En 2015, tanto en el conjunto de España como en Andalucía, más de la mitad de los desempleados llevaban más de 12 meses sin al menos un día de trabajo y, por tanto, con una desconexión del mundo laboral muy prolongada.

Con la finalidad de hacer frente a las consecuencias de esta grave crisis, la mayoría de las regiones españolas pusieron en marcha medidas de empleo y acciones urgentes para tratar de activar a los desempleados, especialmente en aquellos sectores de población con más dificultades de acceso al empleo o con más barreras para la

inserción laboral. Uno de los tipos de PAE que se desarrollaron en algunas CC. AA., tales como Madrid, Cataluña, Aragón, Navarra, Extremadura, Canarias y Andalucía, fueron las iniciativas locales de empleo. En el caso concreto de la Comunidad Autónoma de Andalucía, el Gobierno regional dictó el Decreto-ley 2/2015, de 3 de marzo, de medidas urgentes para favorecer la inserción laboral, la estabilidad en el empleo, el retorno del talento y el fomento del trabajo autónomo, del que se derivaron los siguientes programas[1]: Emple@Joven y programa Emple@30+. Ambos programas ponían énfasis en la creación de empleo en la Comunidad Autónoma de Andalucía, así como en la mejora de la empleabilidad de determinados colectivos con dificultades de acceso al empleo. Una característica distintiva de estos programas de empleo público que favoreció su ejecución frente a otro tipo de PAE, fue el alto nivel de descentralización en la aplicación del programa. Las ofertas de empleo se generan a nivel municipal y esto permitía que estas se adaptasen a las necesidades locales de la población desempleada.

Los objetivos fundamentales de este trabajo se concretan en, primero, caracterizar a los beneficiarios de este programa de Cooperación Social y Comunitaria desarrollado por Andalucía durante los años 2016-2018 y comprobar si, efectivamente, se ha beneficiado a los colectivos más vulnerables y, por ende, con mayores dificultades para su inserción laboral. Segundo, comprobar la empleabilidad posterior de los participantes y su relación de causalidad con ser beneficiarios de dicho programa. Nuestro objetivo final es ofrecer unos resultados que sirvan de referencia para propiciar un debate sobre el diseño de este tipo de políticas activas de empleo. De este debate deberían surgir propuestas de mejora en el diseño de nuevos programas con objetivos similares.

La evaluación de impacto es un método de gran utilidad para responder a la pregunta de en qué medida el cambio en una variable determinada, que mide el resultado de una intervención, es debido a la intervención analizada. Una rigurosa cuantificación de los impactos de los programas o intervenciones implica la elaboración de contrafactuales que permitan identificar y/o estimar qué habría sucedido en ausencia de una intervención específica. El ejercicio de evaluación lo realizaremos aplicando modelos de emparejamiento *(propensity score matching*, PSM). Disponemos de una base de datos de demandantes de empleo proporcionada por el Servicio Andaluz de Empleo (SAE). En dicha base de datos, disponemos de la información necesaria para construir los grupos de tratamiento y control. Los primeros son los participantes en el programa y los segundos son demandantes de empleo que cumplen los requisitos para participar, pero no lo han hecho.

Nuestros resultados fundamentales indican que las iniciativas locales de empleo no han favorecido, en media, un aumento en la probabilidad de salir del paro al empleo. Sin embargo, sí encontramos que para algunos colectivos existen efectos positivos, aunque de pequeña cuantía. Estos colectivos son: demandantes de empleo

[1] Anteriormente, se había ejecutado un programa parecido: Emple@Joven e Iniciativa @mprende+ (Decreto-ley 6/2014, de 29 de abril) y programa Emple@30+ (Decreto-ley 9/2014, de 15 de julio).

mayores de 30 años sin experiencia laboral o con duraciones de paro superiores a los dos años. Por tanto, parece que esta iniciativa puede estar más recomendada para dotar de experiencia laboral a los desempleados, o bien que no hayan trabajado nunca, o bien para parados de larga duración que necesitan reconectar con el sector productivo.

La estructura del resto del artículo es la siguiente. En la segunda sección se presenta una revisión de la literatura empírica. En la tercera sección se describe el programa objeto de análisis. En la cuarta y quinta sección se presentan la metodología empleada y la descripción de la fuente de información y variables utilizadas. En la sexta sección se comentan los resultados encontrados y se finaliza con las conclusiones.

2. Revisión literatura empírica

Es bastante escasa la literatura empírica que evalúa iniciativas de creación directa de empleo público dirigidas a colectivos con dificultades de inserción laboral. Esto es así tanto para España como para otros países europeos. La mayoría de la literatura empírica existente en España se centra en la evaluación de otro tipo de PAE como programas de formación (Blázquez, Herrarte & Sáez, 2019[2]; Cueto & Mato, 2009[3]; Arellano, 2010[4]; INAEM, 2015[5]) o subsidios salariales a la contratación (García-Pérez & Rebollo-Sanz, 2009a), con el objetivo de conocer su impacto sobre los niveles de empleabilidad o estabilidad laboral de los beneficiarios de los programas[6]. Generalmente, a su vez, este tipo de trabajos evalúan los efectos de corto plazo y en pocas ocasiones los efectos de medio plazo relacionados con la calidad y la duración del nuevo empleo (una excepción es García-Pérez & Rebollo-Sanz, 2009b).

---

[2] Estos autores analizan la influencia de programas de formación y asistencia en la búsqueda de empleo en la Comunidad de Madrid sobre la probabilidad de conseguir un empleo para los parados de larga duración (PLD). Los resultados del estudio mostraban que la participación en ambos programas ejercía un efecto positivo en la empleabilidad de los beneficiarios, siendo el efecto de la participación en programas de formación mucho más intenso que el de asistencia a la búsqueda de empleo, tanto en términos de empleabilidad como de calidad en el empleo.

[3] En este trabajo se estudió el efecto de un programa de formación regional voluntario sobre la probabilidad de empleo y encontraron que la formación aumentaba las probabilidades de empleo entre un 8 % y un 9 %.

[4] En él se evalúa el efecto causal de los cursos de formación sobre la duración del desempleo constatando que los cursos de nivel medio redujeron el desempleo entre los trabajadores que lo recibieron.

[5] En 2015 se realizó una evaluación de las PAE llevadas a cabo por el Instituto Aragonés de Empleo (INAEM). En este caso concreto, se analizaron los efectos individuales de tres tipos de medidas (promoción, intermediación y formación). Respecto a las políticas de promoción de empleo se obtuvo que en el primer año un individuo beneficiario de la ayuda trabajaba más tiempo que otro similar que no ha participado. En cuanto a los PLD, grupo objetivo de estas políticas, obtienen un diferencial relativamente más elevado (hasta el tercer mes) que el del resto de desempleados. Respecto a las políticas de formación, los autores encuentran que, en media, estas contribuyen positivamente tanto a la probabilidad de salida del paro (evaluada tras 12 meses) como a la estabilidad del nuevo empleo. No obstante, este efecto positivo no se encuentra para los PLD.

[6] Se han seleccionado algunos de los artículos publicados en los últimos 15 años.

Una referencia fundamental en la literatura empírica relacionada con la evaluación de impacto de las PAE son los metaanálisis de Card, Kluve y Weber (2010, 2018) y Kluve (2010). En estos trabajos se concluye que los programas de creación directa de empleo por parte del sector público son menos eficaces que otro tipo de medidas, como los programas de formación o los de asistencia para la búsqueda de empleo en la mejora de la empleabilidad de los beneficiarios de dichos programas. En concreto, estos metaanálisis constatan que, en media, los programas de empleo público ofrecen malos rendimientos en términos de la empleabilidad de sus beneficiarios. Uno de los mecanismos que pueden generar estos efectos negativos es la duración prolongada de los empleos subvencionados (en la literatura, este efecto se denomina *lock in effect*). La idea fundamental de este enfoque es que el individuo queda «atrapado» en el empleo público cuando la duración del programa es prolongada. A medida que permanecen en el empleo subvencionado, los beneficiarios van recibiendo menos ofertas de trabajo o, si las reciben, las van a rechazar con una alta probabilidad. Esto puede terminar perjudicando las perspectivas laborales al futuro del participante, a pesar de haber ganado algo de experiencia laboral por haber participado en el programa.

Otros trabajos que también hacen evaluación de impacto y llegan a conclusiones similares son los de Kuddo (2009) y Escudero (2016, 2019). Kuddo (2009) revisa las PAE en países de Europa del Este y Asia Central y destaca sus efectos favorables sobre la reducción de la tasa de pobreza en el corto plazo, aunque no encuentra efectos positivos sobre mejoras en el nivel de empleabilidad de los colectivos beneficiarios del programa[7]. En Escudero *et al*. (2019) se analizan los resultados de evaluaciones de impacto de diferentes PAE en América Latina y el Caribe. La evaluación del programa Construyendo Perú llevada a cabo por Escudero (2016) muestra que la iniciativa permitió generar empleo y reducir la inactividad de los grupos beneficiarios, pero a costa de asumir empleos de menor calidad. La evaluación del programa PLANE implementado en Bolivia muestra que, en el contexto de alta rigidez del mercado laboral en el que se desarrolló la política, esta no tuvo impacto en la probabilidad de empleo posterior a la intervención (Hernani-Limarino *et al*., 2011).

Por otra parte, también hay trabajos que apuntan mejoras en las habilidades o en la empleabilidad de los individuos beneficiarios de programas de creación directa de empleo público. Por ejemplo, Roy y Wong (2000) revisa la relación coste-efectividad de programas de creación directa de empleo llevados a cabo por el Gobierno canadiense en las décadas de los años ochenta y noventa del pasado siglo y concluye que este tipo de programas es adecuado para colectivos desfavorecidos porque favorece el desarrollo de habilidades. Hohmeyer y Wolff (2010) evalúan el impacto de tres programas de creación de empleo directo dirigidos a individuos con dificultades de

[7] Kuddo (2009) destaca el caso de un programa masivo público implantado en Bulgaria en 2002, From Social Assistance to Employment, dirigido a grupos desfavorecidos, como aquellos con niveles de educación bajo, desempleados de larga duración y beneficiarios de asistencia social. Los resultados del análisis de impacto indicaban que solo el 8 % de los participantes de estos programas encontraron empleo una vez finalizado, mientras que para el grupo de control este porcentaje era del 16 %.

inserción laboral (beneficiarios de la asistencia social en Alemania) que tuvieron lugar en Alemania entre 2005 y 2008. Los autores concluyen que estos programas contribuyeron a un mejor desempeño laboral entre los participantes.

Para el contexto de España, la literatura empírica centrada en la evaluación de impacto de programas de empleo público es notoriamente escasa, a pesar de que, en los últimos años y tras la Gran Depresión, existen multitud de iniciativas locales de empleo. Entre las excepciones se encuentra el trabajo de Ramos, Surinach y Artís (2009). Estos autores estudian la eficacia de una amplia gama de PAE aplicadas en Cataluña en los años previos a la Gran Depresión (2007-2008). Entre las políticas evaluadas se encuentran las acciones de fomento al empleo en el ámbito local como los Planes de Empleo[8]. Para la mayoría de los programas analizados los autores encuentran un aumento en la probabilidad de empleo de los participantes dos años después del programa, sobre todo para los beneficiarios de planes públicos de empleo[9], apoyo al empleo personalizado y formación profesional.

En resumen, de la revisión de la literatura empírica se puede concluir que existen aún pocos estudios que evalúen la efectividad de las iniciativas locales de empleo. Aún más escasos son los estudios de este tipo de PAE para el caso español. La evidencia encontrada apunta que no suelen tener un marcado efecto positivo en términos de empleabilidad, sobre todo los relacionados con la creación de empleo público. No obstante, la mayoría de los trabajos de evaluación de impacto de las PAE también apuntan la existencia de una gran heterogeneidad en la efectividad de estos programas dependiendo del colectivo al que van dirigidos. Esta gran heterogeneidad de resultados ha puesto de manifiesto la necesidad de que cada programa de empleo público sea diseñado y evaluado en el contexto concreto –mercado de trabajo local– y para el colectivo específico para el que está dirigido (Gerfin & Lechner, 2002; Card, Kluve & Weber, 2018). De hecho, en la actualidad, y a medida que proliferan los trabajos de evaluación causal de las PAE, se considera que un buen diseño de una PAE debe ser aquel que sea multidimensional (incluya formación, asistencia en el proceso de búsqueda de empleo, etc.), flexible y se adapte a las necesidades específicas de cada beneficiario potencial. Es decir, un buen diseño de PAE debe llevar asociado un trabajo previo de perfilado que permita identificar a nivel individual las posibilidades de empleabilidad del desempleado, así como sus necesidades de asistencia en la inserción laboral (Felgueroso, García-Pérez & Jiménez, 2018; Rebollo-Sanz, 2018).

Tampoco los análisis de impacto suelen tener en cuenta otro tipo de efectos relacionados con la incidencia de las PAE sobre la tasa de actividad y/o la tasa de

[8] Programas de contratación laboral de personas que se encuentran en situación de paro para la realización de trabajos de interés público y social en el ámbito local (limpieza, mantenimiento y rehabilitación de edificios municipales, apoyo administrativo y de atención a la ciudadanía, entre otros).

[9] En cuanto al efecto de los diferentes programas, este varía entre el 17,9 % de los Planes de Ocupación y el 4,4 % de las Acciones Integradas (efecto positivo y significativo, para el caso de Planes de Empleo). Para los participantes en los Planes del Empleo, el 63,39 % es el porcentaje de individuos que tenían trabajo a finales de 2007/comienzos de 2008. Del análisis del contrafactual se extrae que si esos individuos no hubiesen participado en dicha iniciativa el porcentaje se reduciría al 45,49 % (caída bastante elevada, no se analizan posibles causas en el trabajo).

pobreza. Nótese que un potencial objetivo de las PAE dirigidas a los colectivos con especiales dificultades de inserción laboral es también reducir la incidencia del denominado «efecto del trabajador desanimado», esto es, evitar el flujo de salida del paro a la inactividad[10].

3. Descripción del programa: las iniciativas locales de empleo

Las PAE se entienden como el conjunto de servicios y programas de orientación, empleo y formación que tienen como objetivo mejorar las posibilidades de acceso al empleo de los desempleados, el mantenimiento del empleo de los que ya lo tienen, así como fomentar el emprendimiento. Son, por tanto, un conjunto de medidas e iniciativas que buscan mejorar el funcionamiento del mercado laboral mediante instrumentos que ayuden a la creación y el mantenimiento del empleo. Entre los distintos tipos de PAE, las que se evalúan en este trabajo se engloban dentro de las medidas de creación directa de empleo en el sector público (*i. e.* empleo en las distintas Administraciones públicas o empresas que producen bienes o servicios públicos). Dichos programas, en general, van dirigidos a personas con dificultades de inserción laboral con el fin de mantenerlos en contacto con el mercado de trabajo.

Concretamente, la Ley andaluza 2/2015, de 29 de diciembre, de medidas urgentes para favorecer la inserción laboral, la estabilidad en el empleo, el retorno del talento y el fomento del trabajo autónomo, regula un conjunto de programas cuyo objetivo principal es dar respuesta a las necesidades urgentes de empleo de la población desempleada andaluza. Dentro de estos programas se incluía la denominada Iniciativa Cooperación Social y Comunitaria Emple@Joven y Emple@30+, cuya dotación presupuestaria fue de 250 millones de euros[11]. El programa Emple@Joven estaba dirigido a jóvenes de entre 16-29 años y el programa Emple@30+ a personas de entre 30 y 65 años[12]. Estos programas tenían por objeto mejorar la empleabilidad de las personas demandantes de empleo que se encontraban con dificultades de acceso al mismo, posibilitando la adquisición de competencias profesionales a través de nuevas experiencias de empleo.

---

[10] Este efecto hace referencia al desaliento de las personas que desean trabajar y no encuentran un empleo en periodos de crisis. Este efecto suele estar relacionado con el estancamiento de la actividad económica y las negativas perspectivas en materia de creación de empleo y se agudiza a medida que el individuo permanece desempleado y sus posibilidades de empleo se reducen.

[11] Otras medidas incluidas en la Ley 2/2015, de 29 de diciembre, de medidas urgentes para favorecer la inserción laboral, la estabilidad en el empleo, el retorno del talento y el fomento del trabajo autónomo son: *i*) Becas para el desarrollo de prácticas profesionales en empresas, con una dotación de dos millones de euros. *ii*) Ayudas para la contratación de personas titulares del Bono de Empleo Joven, con una dotación de 70 millones de euros. *iii*) Iniciativa de Proyectos de Interés General y Social generadores de empleo, con una dotación de 15 millones de euros. *iv*) Incentivos a la contratación indefinida de personas jóvenes, con una dotación de casi 48 millones de euros. *v*) Ayudas al trabajo autónomo, con una dotación de 17 millones de euros. Las iniciativas recogidas en esta ley están cofinanciadas con Fondo Social Europeo.

[12] La dotación para cada programa fue de 150 millones para la contratación de los jóvenes entre 18 y 29 años y 100 millones para la contratación del colectivo de más de 30 años.

En este tipo de programas, las ayudas eran concedidas a los ayuntamientos para la contratación de personas desempleadas residentes en el municipio. Por su parte, las personas destinatarias finales de la iniciativa eran el conjunto de demandantes de empleo no ocupados que reunían una serie de requisitos que comentaremos más adelante. Esta línea de ayudas era gestionada en régimen de concurrencia no competitiva, por lo que los ayuntamientos solo tenían que presentar la solicitud en los términos especificados en la convocatoria para que esta se resolviese favorablemente. El importe de la ayuda consistía en una cantidad a tanto alzado que se determinaba atendiendo a la duración del contrato y al grupo de cotización a la Seguridad Social del individuo contratado[13]. La cuantía máxima de la ayuda a percibir por los municipios se determinaba en función del nivel de desempleo y la población de cada municipio, distinguiendo los límites de edad antes mencionados.

El número total de individuos beneficiarios de estas iniciativas locales de empleo dependían, en primera instancia, de las solicitudes de los ayuntamientos, puesto que sobre ellos recaía la responsabilidad de presentar una oferta de empleo ante el Servicio Andaluz de Empleo (SAE). Este aspecto del diseño del programa (la oferta de trabajo la diseña el ayuntamiento), se ha considerado una característica distintiva y ventajosa de las iniciativas locales de empleo. Esto es así porque favorecía que la oferta de empleo creada por el programa estuviese ligada a las necesidades de empleo local tanto desde el punto de vista de la demanda (tipo de vacantes locales), como desde el punto de vista de la oferta (perfiles de desempleados del municipio). Una vez presentada la oferta por parte del ayuntamiento, el SAE debía tratar de proporcionar tres personas candidatas adecuadas y disponibles en base al perfil solicitado. En caso de que hubiese más de tres participantes potenciales, la convocatoria establecía un orden de prelación según los siguientes parámetros: *i*) personas beneficiarias del Ingreso Mínimo de Solidaridad (IMS, en adelante); *ii*) desempleados de larga duración (PLD) que hayan agotado la prestación por desempleo de nivel contributivo o asistencial en los últimos doce meses; *iii*) PLD en general; *iv*) personas desempleadas en general. Por tanto, el programa articulaba la contratación en base a una definición de vulnerabilidad laboral sustentada en las variables renta (a través de la percepción o no del IMS) y duración del desempleo (a través de la pertenencia o no al colectivo de PLD)[14].

La oferta de trabajo debía estar formulada de forma precisa y ajustada a los requerimientos del puesto de trabajo. Para ello, los ayuntamientos debían tomar como referencia las categorías profesionales y los criterios de realización asociados a alguna unidad de competencia incluida en el Catálogo Nacional de Cualificaciones Profesionales (CNCP) vigentes para definir la oferta de trabajo. Todo ello, con objeto de que la experiencia profesional adquirida en el desempeño del puesto de trabajo se

[13] Estas son: grupo 1: 1.700 euros al mes; grupo 2: 1.600 euros al mes; grupo 3: 1.500 euros al mes; grupos 4-10: 1.300 euros al mes.

[14] Adicionalmente, dentro de cada grupo de prioridad se ordenarán los candidatos atendiendo a la mayor disponibilidad para el empleo, la fecha de solicitud de ocupación o, en su defecto, a la fecha de inscripción, de la más antigua a la más reciente.

pudiese acreditar *a posteriori*. En este sentido, esta iniciativa introdujo un aspecto novedoso respecto a sus convocatorias precedentes: el proceso de tutorización de las personas contratadas[15]. Es decir, a la finalización del período de contratación, el ayuntamiento debía elaborar y entregar un certificado individual a cada participante en el que quedase constancia de las competencias profesionales adquiridas. Finalmente, se debía realizar un informe de seguimiento global del Proyecto de Cooperación Social y Comunitaria en el que se reflejaran los resultados obtenidos. Este documento debía ser entregado al SAE junto con la justificación económica de la ayuda[16].

En relación con el tipo de contrato que debía usar el ayuntamiento, el programa definió la modalidad de contrato de obra o servicio determinado. La duración del contrato dependía del programa. En el programa Emple@30+ se estableció un período mínimo de tres meses, debiendo formalizarse por meses completos, hasta un máximo de seis. En el programa Emple@Joven, la duración mínima establecida fue de seis meses. Además, los contratos se tenían que concertar a jornada completa.

En cuanto a los requisitos y criterios de las personas participantes, estas tenían que estar inscritas como demandantes de empleo no ocupadas y ser residentes en el municipio de referencia. En el programa Emple@30+ también se exigía que el participante estuviese inscrito en el fichero del Sistema Nacional de Garantía Juvenil.

3.1. Perfil de los contratados (grupo de tratamiento)

La base de datos que se usa para realizar la evaluación de impacto procede del SAE, tanto la relativa a los participantes del programa como al grupo de control que se describirá en la siguiente sección. Recordemos que, dado el funcionamiento del programa, el SAE tiene toda la información estadística de los demandantes de empleo que se han beneficiado de las iniciativas locales de empleo. El número global de personas contratadas por los ayuntamientos a julio de 2018 ascendió a un total de 33.295 personas, de las cuales 17.770 eran participantes en el programa empleo joven y 15.525 participantes en el programa 30+.

Comenzamos presentando en el Cuadro 1 la composición de los beneficiarios de las iniciativas locales de empleo para cada uno de los programas según un conjunto de características socioeconómicas. Dichas características se concretan en el siguiente conjunto de variables observables: género (mujer), edad, niveles de estudio (tres niveles de estudio: educación primaria, secundaria y postsecundaria), nacionalidad (inmigrante frente a nativo), beneficiario del IMS (IMS, variable ficticia que toma valor la unidad si es beneficiario y cero en caso contrario), duración del paro

[15] El proceso de tutorización consiste fundamentalmente en cumplimentar el cuaderno de seguimiento, donde se debe detallar las realizaciones profesionales incluidas en cualificaciones vigentes del CNCP al objeto de que se pueda realizar un seguimiento de las competencias adquiridas con la práctica laboral.

[16] A las obligaciones anteriores, habría que sumar obligaciones relativas a la contabilidad de los gastos, ejecución de la obra o servicio, y obligaciones de información y publicidad establecidas por la normativa comunitaria y el Servicio Andaluz de Empleo, entre otras que se mencionan en la convocatoria.

CUADRO 1
CARACTERÍSTICAS SOCIOECONÓMICAS DE LOS PARTICIPANTES
(En %)

		30+ Media (sd)	Joven Media (sd)
	Mujer (%)	46 (0,50)	51 (0,50)
	Edad (años)	44,5 (7,9)	24,1 (3,4)
	Educación primaria	9 (0,28)	6 (0,39)
	Educación secundaria	57 (0,50)	65 (0,48)
	Educación postsecundaria	35 (0,48)	37 (0,48)
	Inmigrante	2 (0,13)	1 (0,10)
	Beneficiario IMS	22 (0,41)	7 (0,26)
	Minusválido	4 (0,18)	4 (0,20)
	Duración del paro (meses)	13,9 (9,1)	12,2 (8,7)
	12-23 meses	20 (0,40)	25 (0,43)
	>= 24 meses	35 (0,48)	23 (0,42)
	Sin experiencia laboral	11 (0,31)	35 (0,48)
	Meses experiencia laboral	47,8 (61,8)	5,3 (10,7)
	Colectivo CDIL	69 (0,46)	63 (0,48)
	Tratado L2014	11 (0,31)	19 (0,39)
Tamaño municipio	< 5 mil	22 (0,41)	14 (0,35)
	5-10 mil	14 (0,34)	11 (0,32)
	10-20 mil	18 (0,38)	16 (0,37)
	20-50 mil	18 (0,38)	20 (0,40)
	50-100 mil	11 (0,32)	14 (0,34)
	> 100 mil	18 (0,38)	24 (0,43)
Provincia	Almería	8 (0,27)	8 (0,27)
	Cádiz	13 (0,34)	16 (0,37)
	Córdoba	11 (0,31)	11 (0,31)
	Granada	13 (0,33)	11 (0,32)
	Huelva	7 (0,27)	7 (0,25)
	Jaén	11 (0,31)	9 (0,29)
	Málaga	14 (0,34)	14 (0,35)
	Sevilla	23 (0,42)	23 (0,42)
Total		**17.770**	**15.525**

FUENTE: Elaboración propia con los datos procedentes del SAE.

(meses), experiencia laboral, beneficiario de una iniciativa local de empleo anterior (Tratado L2014), tamaño del municipio (seis variables ficticias de tamaño de municipio según número de habitantes) y provincia de residencia. También construimos una variable *Colectivo CDIL*, que toma valor la unidad si se trata de individuos que por sus características demográficas se pueden catalogar de vulnerables y, por tanto, con dificultades de inserción laboral, y cero en caso contrario. Dichos colectivos son PLD (duración del episodio de paro superior a 12 meses), individuos con estudios primarios, inmigrantes o perceptores del IMS[17].

El Cuadro 1 muestra que el programa cumple uno de sus objetivos fundamentales[18], este es, dirigirse a *Colectivos CDIL*. Este colectivo representa más del 60 % de los beneficiarios en ambos programas (el 69 % en los mayores de 30 y del 63 % en los jóvenes). Como era de esperar, entre los participantes hay un alto porcentaje de PLD (el 55 % en el caso del programa de mayores de 30 años y el 48 % en el caso del programa para jóvenes). En el programa empleo 30+ los beneficiarios del IMS representan el 22 %. En el programa empleo joven, ese porcentaje no es tan elevado porque para ser beneficiario del IMS se debe tener cargas familiares y hay muchos jóvenes que naturalmente no se encuentran en esa situación. Por tanto, respecto al potencial efecto de estos programas, la composición de la muestra de tratados tiene una lectura muy positiva.

3.2. Descripción estadística del programa: características del programa

En el Cuadro 2 se presenta una estadística básica del tipo de empleo realizado por el individuo beneficiario del programa. El tipo de empleo se caracteriza a partir de las siguientes variables: tipo de contrato (obra y servicio, eventual), duración del contrato, tipo de ocupación y año del contrato subvencionado. La duración del episodio de empleo se describe con cuatro tramos: entre 1 y 3 meses, entre 4 y 6 meses, entre 7 y 12 meses y más de 12 meses. El tipo de ocupación se agrupa en tres categorías principales: alta (grupos de cotización 1-3), media (grupos de cotización 4-7) y baja (grupos de cotización 8 y 9). Finalmente, para enmarcar temporalmente el análisis, los episodios de empleo hacen referencia al periodo que transcurre desde el comienzo del programa, este es noviembre de 2016, hasta agosto de 2018.

La duración observada de los contratos está íntimamente ligada a los periodos mínimos de contratación establecidos por la normativa reguladora para cada uno de los programas (6 meses para el programa joven y 3 meses para el 30+). Si analizamos el periodo de duración de las contrataciones con más detenimiento, se observa

[17] Dentro de esta base de datos de individuos tratados proporcionada por el SAE tenemos algunas situaciones «anómalas» o poco comunes que se eliminarán de la base de datos objeto de análisis. En concreto, estas son: *i)* individuos con varios tratamientos (=varios contratos bajo el programa), nos quedamos con el último episodio; *ii)* minusválidos; *iii)* VVG; *iv)* individuos con contratos (tratamiento) de duración indeterminada.

[18] El programa define los colectivos prioritarios –es decir, en caso de tener que ordenar a los demandantes de empleo, los trabajadores con estas características tienen mayor probabilidad de beneficiarse del programa–, los cuales son el ser beneficiarios del IMS y ser PLD (> 12 meses).

CUADRO 2
CARACTERÍSTICAS DEL EMPLEO
(En %)

		30+ Media (sd)	Joven Media (sd)
Duración del contrato	1-3 meses	42 (0,5)	4 (0,2)
	4-6 meses	53 (0,5)	87 (0,3)
	7-12 meses	4 (0,2)	8 (0,2)
	> 12 meses	1 (0,0)	0 (0,06)
Tipo de ocupación (cualificación)	Alta	29 (0,4)	32 (0,4)
	Media	43 (0,5)	31 (0,4)
	Baja	28 (0,4)	37 (0,5)
Tipo de contrato	Obra y servicio	95 (0,2)	96 (0,2)
	Eventual y otros	5 (0,2)	3 (0,2)
Año del contrato subvencionado	2016	2 (0,1)	2 (0,1)
	2017	11 (0,2)	9 (0,2)
	2018	87 (0,3)	89 (0,3)
Tipos de ocupación (CNO-1d)	Directores y gerentes	0,1	0,1
	Técnicos y profesionales	19	17
	Técnicos, profesionales de apoyo	9,9	15
	Administrativos, empleos de oficina	5,3	7,9
	Servicios de restauración, vendedores	6,2	5,8
	Cualificado agrícola, ganadero, forestal	5,5	5,9
	Artesanos/trabajadores cualificados industria/construcción	25	11
	Operadores de instalaciones y maquinaria	6,2	0,1
	Ocupación elemental	27	36

FUENTE: Elaboración propia con los datos procedentes del SAE: extracción de la base de datos de demandantes de empleo del SAE.

que en el programa Emple@Joven predominó el periodo mínimo de contratación regulado por la norma (el 87 % de las contrataciones han sido de 6 meses), mientras que en el Emple@30+ las contrataciones se extendieron más allá del periodo mínimo exigido de 3 meses (casi el 58 % han sido contrataciones de más 3 meses, y de estas casi la mitad, el 46,8 %, han tenido una duración de 6 meses). Con relación a la jornada laboral, la mayoría de los contratos fueron a tiempo completo. Solo el 4,1 % de los contratos en el caso de los jóvenes y el 5,9 % en el programa 30+ se realizaron a tiempo parcial.

En las iniciativas locales de empleo, las ofertas de empleo generadas están fundamentalmente vinculadas a las competencias municipales. El carácter descentralizado de este tipo de programas permite que se generen una amplia variedad de tipos de empleo. Esta característica se pone de manifiesto con el análisis por gran grupo de ocupación[19]. La mayor proporción de las contrataciones se realizaron en ocupaciones elementales que requerían una baja especialización[20] (un 27 % en el programa Emple@30+ y 36 % en el programa Emple@Joven). No obstante, hay dos categorías con una alta presencia que sí requiere una especialización, como es el caso de los artesanos y trabajadores cualificados de las industrias manufactureras y la construcción (25 % en el programa Emple@30+), y las ocupaciones de técnicos y profesionales científicos e intelectuales (19 % en el programa Emple@30+ y 17 % en el programa Emple@Joven).

3.3. Descripción estadística del programa: indicadores de resultados

El principal indicador estadístico para medir la eficacia del programa en términos de mejora del nivel de empleabilidad es la probabilidad de salir del paro al empleo. Para evaluar dicha probabilidad, el SAE nos proporciona información sobre la situación laboral de los beneficiarios de las iniciativas locales de empleo seis meses después de haber finalizado el contrato. Es decir, la base de datos que nos proporciona el SAE nos permite identificar si el individuo está o no empleado seis meses después del contrato, así como algunas características del empleo.

Una posible objeción a esta forma de presentar la información es que es posible que un participante sea clasificado como no empleado tras el programa si ha tenido una experiencia de empleo corta que ha finalizado antes del momento de la recogida de la información de empleabilidad por parte del SAE[21].

Para medir la eficacia del programa a medio plazo, tratamos también de evaluar la estabilidad del nuevo empleo y/o calidad del mismo. Con este fin se analizan algunas características del nuevo empleo tales como el tipo de contrato (indefinido frente a temporal), la jornada laboral (tiempo completo frente a tiempo parcial), tipo de cualificación en el empleo (medida por el grupo de tarifa) o duración del contrato. Con esta información se puede evaluar, por ejemplo, si el programa ha favorecido cierta estabilidad laboral (*i. e.* el individuo tratado ha conseguido un contrato indefinido tras su participación en el programa) o ha favorecido mejoras en la cualificación[22] .

---

[19] El tipo de ocupación se basa en la Clasificación Nacional de Ocupaciones a un dígito.

[20] Entre los diferentes tipos de empleo, ocupa un lugar predominante la reparación y mantenimiento de calles, jardines y otros espacios/edificios públicos.

[21] Dado que el objetivo del programa es medir la empleabilidad del individuo a medio plazo (más que a corto plazo), esta forma de recoger la información es apropiada y no debería introducir sesgos importantes en el análisis de los resultados.

[22] Otro indicador interesante sería la duración del nuevo empleo, sobre todo si es de duración superior a un tiempo mínimo.

En el Cuadro 3 se presentan los principales indicadores de eficiencia del programa. Estos son la probabilidad de salida del paro y el grado de estabilidad del nuevo empleo según los indicadores arriba mencionados. De los 33.295 beneficiarios del programa 30+, un total de 9.801 han obtenido un contrato en los seis meses posteriores a la finalización del programa, lo que depara una *tasa de contratación posterior al programa* del 29,4 %. En el caso de los beneficiarios del programa empleo joven, los resultados son aún más favorables y el 33 % de los participantes tienen un contrato seis meses después del programa. Dicho de otro modo, casi 1 de cada 3 participantes en ambos programas estaba empleado a los seis meses de finalización del mismo.

De los que consiguen un empleo, la mayoría lo hacen con un contrato temporal (97-98 % dependiendo del programa). Esto se corresponde con la dinámica propia del mercado de trabajo español, donde más del 95 % de los nuevos contratos son temporales. En relación con la jornada laboral, un 47 % tienen un empleo a tiempo parcial en el programa Emple@30+, mientras que un 60 % tienen un empleo a tiempo parcial tras el programa Emple@Joven. Por tanto, la incidencia en cuanto a parcialidad es muy alta, sobre todo en el colectivo de los jóvenes[23]. La última medida de empleabilidad presentada tiene que ver con la eventual o*cupación posterior* en los seis meses posteriores a la finalización del programa. La mayoría de las ocupaciones están en los niveles bajos (48 % para el programa Emple@30+) o medios (43 % para el programa Emple@Joven).

CUADRO 3
INDICADORES DE INSERCIÓN LABORAL TRAS EL PROGRAMA
(En %)

		30+ Media (sd)	**Joven Media (sd)**
Sale del paro al empleo		29 (0,45)	33 (0,47)
Condicionado a salir del paro	Contrato temporal	98 (0,14)	97 (0,18)
	Tiempo parcial	47 (0,50)	60 (0,49)
Duración del contrato	< 3 meses	34 (0,46)	34 (0,45)
	3-6 meses	6 (0,28)	9 (0,28)
	Con fecha de finalización	45 (0,50)	55 (0,50)
Tipo de ocupación	Alta	13 (0,34)	20 (0,40)
	Media	39 (0,49)	43 (0,50)
	Baja	48 (0,50)	37 (0,48)

FUENTE: Elaboración propia con los datos procedentes del SAE: extracción de la base de datos de demandantes de empleo del SAE.

[23] Según la EPA, la tasa de empleo a tiempo parcial en Andalucía en esos años se sitúa en torno al 16,2 % (14,8 % para el caso de España).

4. Descripción de la metodología de la evaluación de impacto: efecto medio del tratamiento

Si bien el uso de la estadística descriptiva presentada en la sección anterior puede ser indicativa de ciertas bondades de esta iniciativa, no es posible extraer conclusiones robustas sobre la efectividad del programa sobre los niveles de empleabilidad de los beneficiarios. Para ello, es necesario realizar un análisis contrafactual que permita testar la existencia de una relación causal entre la participación en el programa y la ganancia en empleabilidad posterior de los participantes. Es decir, hasta que no se demuestre que los resultados de inserción laboral de individuos iguales a los tratados, pero que no se han beneficiado del programa, son peores a los de los tratados. Esta es a la pregunta que se trata de contestar en el ejercicio de evaluación causal: ¿qué le hubiese pasado –en términos de duración total del episodio de paro, probabilidad de encontrar un empleo o probabilidad de engrosar las listas de inactivos– a los desempleados participantes de las iniciativas locales de empleo (Y_{1i}) si no hubieran participado en la misma? Esta última situación es lo que se denomina, resultado «contrafactual« (Y_{0i}).

Para poder valorar la diferencia entre el resultado observado (Y_{1i}) y el «contrafactual» (Y_{0i}), la aproximación más habitual consiste en comparar la situación de dos grupos de individuos, uno de los cuales se ha beneficiado de la política (conocido como grupo de tratamiento o simplemente tratados, Y_{1i}) y el otro no (conocido como grupo de control o simplemente controles, Y_{0i}). Es decir, el efecto del tratamiento en el individuo se calcularía como la diferencia entre los dos resultados ($Y_{1i} - Y_{0i}$). La naturaleza de estos estudios es del tipo cuasiexperimental pues, en la práctica, los individuos no son asignados de manera aleatoria a los grupos de tratamiento y control, sino que existen unos criterios de elegibilidad que determinan dicha asignación. En el caso del programa que evaluamos, la selección de los participantes depende del carácter más o menos vulnerable del individuo sobre la base de las variables renta –a través de la percepción o no del IMS– y desempleo –a través de la pertenencia al colectivo de PLD.

En este trabajo, como metodología de evaluación de impacto, utilizamos la técnica conocida como método de emparejamiento basado en la probabilidad de asignación al tratamiento (*propensity score matching,* PSM)[24]. El método de emparejamiento consiste en hacer parejas (*matches*) entre personas del grupo de tratamiento y personas del grupo de control con algunas características de partida similares, para después comparar sus resultados finales y, posteriormente, estimar el efecto medio

[24] Se utiliza esta técnica al no ser posible realizar otro tipo de análisis en condiciones experimentales que permitirían distribuir los «tratamientos» de forma aleatoria y estimar los efectos netos de manera más precisa. No obstante, este método permite simular esas condiciones mediante un diseño cuasiexperimental (o de experimento natural) para aproximarse más al efecto neto real de la intervención. Asimismo, se elige este procedimiento frente a otros análisis cuasiexperimentales por la falta de una regla clara y repentina de aplicación de la intervención (diseño de regresión discontinua) o de una evolución similar previa de dos grupos que luego puedan divergir a partir del tratamiento (diferencias-en-diferencias), entre otros aspectos.

del tratamiento *(average treatment effect)*. El aspecto clave del procedimiento de emparejamiento consiste en un algoritmo que permite identificar, con la máxima precisión posible, aquellos individuos no tratados parecidos a cada uno de los participantes en el programa. Este procedimiento de emparejamiento se basa en un modelo econométrico que estima la probabilidad de ser beneficiario del programa, a partir de la muestra de tratados y no tratados, y usando como variables explicativas de dicha probabilidad un conjunto de características observables.

A partir de dicho algoritmo se identifica una muestra de individuos contrafactuales, que en base a la probabilidad estimada de ser tratado –misma propensión a participar en el programa–, es casi igual al grupo de individuos tratados (Smith & Todd, 2005). Al emparejar a los individuos tratados con contrafactuales con la «misma o parecida» probabilidad de participar en el programa, se evita el problema del sesgo de selección: es decir, el hecho de que las personas que participaron en el programa fueran diferentes en sus niveles de inserción laboral potencial frente al conjunto de los individuos sin trabajo. Una vez se ha hecho el «emparejamiento» de tratados y controles, se comparan los resultados del grupo «tratado» en este caso de participantes frente a los de los casos «de control» y se estima el efecto medio del tratamiento sobre los tratados (*average treatment effect on the treated,* ATET).

5. Radiografía de los grupos de tratamiento y control

Cuando hemos descrito el programa, hemos hecho una caracterización inicial de los participantes de las iniciativas locales del empleo. En esta sección presentamos un análisis comparativo entre la muestra del grupo de tratamiento y el grupo de control, con la finalidad de observar si el perfil de participantes en cada uno de los programas Emple@30+ y Emple@Joven responde al patrón de desempleo que caracteriza a la región andaluza.

La estadística presentada en la sección 3 parece indicar que las iniciativas de empleo evaluadas están orientadas de forma efectiva a los colectivos más desfavorecidos, siendo estos los conformados por personas desempleadas con menor probabilidad de encontrar un trabajo en el medio plazo. Como se ha descrito en el punto anterior, la convocatoria de las iniciativas locales de empleo precisa unos criterios de selección que ordena a los potenciales beneficiarios de más a menos vulnerable. Por tanto, la participación en estos programas no es aleatoria y se basa en un conjunto de variables observables. Esto implica que una mera comparación de los resultados de los individuos beneficiarios y no beneficiarios del programa puede llevar a conclusiones erróneas dadas las especiales características del colectivo de beneficiarios.

La base de datos de partida para construir la muestra de individuos que componen el grupo de control la construye el SAE a partir de los demandantes de empleo registrados. Más concretamente, el SAE realiza una extracción aleatoria de los demandantes de empleo registrados a fecha de 30/06/2018, excluyendo a los participantes en otros programas de la Ley 2/2015. El SAE establece como fecha de corte para la

extracción de la muestra ese mes, al corresponderse con el mes de mayor volumen de participantes que finalizan el programa, en concreto el 49,8 %. Por tanto, para la contabilización del tiempo de búsqueda de empleo, se considera junio de 2018 como el momento 0 para el grupo de control y la propia finalización del programa para el grupo de tratamiento. En ambos casos, tanto para los beneficiarios del programa –tratado–, como para los no beneficiarios –controles– la información socioeconómica usada en este trabajo procede del SAE.

La dimensión de la muestra de individuos que forman parte del grupo de control que nos proporciona el SAE es adecuada (Cuadro 4). Para el programa Emple@30+ disponemos de 134.303 observaciones y para el programa Emple@Joven de 39.981. Esto significa que del total de jóvenes demandantes de empleo que componen la muestra que usamos en el ejercicio de impacto, el 70 % pertenecen al grupo de control y, del total de mayores de 30 años, el 90 % pertenecen al grupo de control. En consecuencia, para cada unidad de tratado habrá, con una alta probabilidad, un número suficiente de individuos en la muestra de control.

En el Cuadro 5 comparamos las características socioeconómicas de los beneficiarios del programa con las del grupo de no tratados. En dicho cuadro presentamos la media –si es una variable continua– o el porcentaje –si es una variable discreta–, para ambos grupos de individuos, así como la significatividad estadística del test de diferencias de medias. Si bien se observa una relativa homogeneidad entre ambos grupos, emergen ciertos aspectos diferenciados de interés, en parte consecuencia de los propios criterios de priorización establecidos por la convocatoria para la participación en el programa. Más concretamente, el análisis estadístico del Cuadro 6 confirma que el grupo de individuos participantes en cada uno de los programas es diferente de forma estadísticamente significativa a la de controles. Dichas diferencias apuntan que el programa se ha dirigido a colectivos con especiales dificultades de inserción laboral dentro del panorama del mercado laboral andaluz.

CUADRO 4
ESTADÍSTICA BÁSICA: COMPOSICIÓN DE LA MUESTRA DEL EJERCICIO DE EVALUACIÓN
(En %)

Programa	Control	Tratado	Totales
30+	134.303	13.819	148.122
	90,6	9,13	100
Joven	39.981	16.480	56.461
	70,8	28	100
Total	**174.284**	**28.771**	**204.583**
	85,2	**14,8**	**100**

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE: extracción de la base de datos de demandantes de empleo del SAE.

CUADRO 5
ESTADÍSTICA DESCRIPTIVA CARACTERÍSTICAS SOCIOECONÓMICAS: DIFERENCIAS ENTRE LOS BENEFICIARIOS Y NO BENEFICIARIOS DEL PROGRAMA
(En %)

	30+			Joven		
	Control Mean (sd)	**Tratado Mean (sd)**	**Diff t-stat**	**Control Mean (sd)**	**Tratado Mean (sd)**	**Diff t-stat**
Mujer	61 (0,49)	47 (0,50)	***	56 (0,50)	51 (0,50)	***
< 20 años	–			13 (0,33)	4 (0,18)	***
20-24 años	–			34 (47)	35 (0,48)	***
25-29 años	–			54 (0,50)	62 (0,49)	***
30-39 años	30 (0,46)	30 (0,46)	–			
40-49 años	35 (0,48)	38 (0,49)	***	–		
55-59 años	17 (0,37)	12 (0,32)	***	–		
Educación primaria	20 (0,40)	9 (0,28)	***	19 (0,39)	6 (0,24)	***
Educación secundaria	66 (0,48)	57 (0,5)	***	65 (0,48)	56 (0,50)	***
Educación superior	15 (0,35)	34 (0,48)	***	16	37 (0,48)	***
Inmigrante	8 (0,2)	2 (0,1)	***	7 (0,2)	1 (0,1)	***
IMS	1 (0,1)	22 (0,4)	***	1 (0,1)	7 (0,2)	***
Minusválido	3 (0,2)	4 (0,1)	***	3 (0,1)	4 (0,2)	***
Duración paro 1-3 meses	24 (0,4)	16 (0,3)	***	45 (0,5)	18 (0,3)	***
4-6 meses	14 (0,3)	13 (0,3)	***	20 (0,4)	16 (0,3)	***
7-11 meses	17 (0,3)	17 (0,3)	***	17 (0,3)	19 (0,3)	***
12-23 meses	15 (0,3)	20 (0,4)	***	10 (0,3)	25 (0,4)	***
>= 24 meses	30 (0,4)	35 (0,4)	***	8 (0,2)	23 (0,4)	***
Sin exp. laboral	27 (0,4)	11 (0,3)	***	39 (0,4)	35 (0,4)	***
Meses exp. laboral	35,5 (59,7)	47,8 (61,8)	***	4,1 (10,7)	5,4 (10,5)	***
Colectivo desfavorable (CDIL)	65 (0,4)	68 (0,4)	***	57 (0,5)	63 (0,4)	***
Tratado L2014	2 (0,1)	11 (0,3)	***	4 (0,1)	19 (0,3)	***
Tamaño municipio < 5 mil	8 (0,2)	21 (0,4)	***	10 (0,3)	14 (0,3)	***
5-10 mil	8 (0,2)	14 (0,3)	***	9 (0,2)	11 (0,3)	***
10-20 mil	12 (0,3)	18 (0,4)	***	13 (0,3)	16 (0,3)	***
20-50 mil	18 (0,3)	18 (0,4)	**	19 (0,3)	20 (0,4)	***
50-100 mil	16 (0,3)	12 (0,3)	***	15 (0,3)	14 (0,3)	***
> 100 mil	38 (0,4)	18 (0,3)	***	35 (0,4)	24 (0,4)	***
Provincia Almería	8 (0,2)	8 (0,2)		8 (0,2)	8 (0,2)	
Cádiz	18 (0,3)	14 (0,3)	***	17 (0,3)	16 (0,3)	*
Córdoba	9 (0,2)	11 (0,3)	***	10 (0,3)	11 (0,3)	***
Granada	10 (0,3)	13 (0,3)	***	11 (0,3)	11 (0,3)	
Huelva	7 (0,2)	8 (0,2)	***	6 (0,2)	7 (0,2)	***
Jaén	6 (0,2)	10 (0,3)	***	8 (0,2)	9 (0,2)	***
Málaga	18 (0,3)	14 (0,3)	***	16 (0,3)	14 (0,3)	***
Sevilla	25 (0,4)	23 (0,4)	***	24 (0,4)	23 (0,4)	***

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE: extracción de la base de datos de demandantes de empleo del SAE.

NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %.

CUADRO 6

ESTADÍSTICA DESCRIPTIVA DE LOS INDICADORES DE INSERCIÓN LABORAL TRAS EL PROGRAMA: DIFERENCIAS ENTRE LOS BENEFICIARIOS Y NO BENEFICIARIOS DEL PROGRAMA

(En %)

	30+			Joven		
	Control Mean (sd)	**Tratado Mean (sd)**	**Diff t-stat**	**Control Mean (sd)**	**Tratado Mean (sd)**	**Diff t-stat**
Probabilidad estar empleado	27 (0,4)	29 (0,4)	***	42 (0,5)	33 (47)	***
Tipo de contrato						
Contrato temporal	96 (0,2)	98 (0,14)	***	96 (0,20)	97 (0,18)	***
Tiempo parcial	56 (0,5)	47 (0,5)	***	64 (0,48)	0,59 (0,49)	***
Duración del contrato						
< 3 meses	36 (0,4)	34 (0,4)	***	38 (0,46)	34 (0,45)	***
3-6 meses	11 (0,3)	6 (0,3)	***	12 (0,39)	9 (0,35)	***
Con fecha finalización	54 (0,5)	45 (0,5)	***	63 (0,48)	55 (0,5)	***
Tipo de ocupación						
Alta	9 (0,3)	13 (0,3)	***	12 (0,33)	20 (0,4)	***
Media	45 (0,5)	39 (0,5)	***	47 (0,50)	43 (0,5)	***
Baja	46 (0,5)	48 (0,5)		41 (0,49)	37 (0,48)	***

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE: extracción de la base de datos de demandantes de empleo del SAE.

NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %.

Así, encontramos que los colectivos con alguna o varias dificultades de inserción laboral representan el 68 % de los beneficiarios, 3 puntos porcentuales (pp, en adelante) por encima del peso relativo de estos colectivos en el grupo de control del programa Emple@30+. Las diferencias son aún mayores en el programa Emple@ Joven, al alcanzar este diferencial 5 pp (63 % para los tratados y 57 % para los controles). En el colectivo de beneficiarios hay un alto porcentaje de PLD (55 % y 48 % en la muestra 30+ y joven, respectivamente) y de beneficiarios del IMS (22 % y 7 % en la muestra 30+ y joven, respectivamente), en comparación con la muestra de no beneficiarios (los PLD representan el 45 % y 18 % en la muestra 30+ y en joven, respectivamente, y los beneficiarios del IMS representan el 1 % en ambos casos). Este dato tiene una lectura positiva respecto al efecto que está teniendo dicha política en cuanto a la integración laboral de los colectivos más desfavorecidos. Estos son precisamente los que suelen presentar más dificultades de acceso al mercado de trabajo y, por tanto, los expuestos a una mayor vulnerabilidad y riesgo de exclusión social.

Sí es cierto que en algunas otras dimensiones, tales como el género o el nivel educativo de los participantes, la composición de la muestra de control es algo diferente a la que se esperaría. Así, en el grupo de tratamiento hay menos mujeres y menos individuos con estudios primarios que en el grupo de control. Se esperaría que en el grupo de tratamiento hubiese más individuos con estudios primarios, desde el punto de vista de que estos pueden ser más vulnerables a la hora de acceder al mercado de trabajo.

6. Resultados de la estimación: evaluación de impacto de las iniciativas locales de empleo

El objetivo del trabajo de evaluación es analizar la eficacia de las iniciativas locales de empleo. A través de un ejercicio de evaluación de impacto, pretendemos determinar si el programa ha supuesto una mejora en los niveles de empleabilidad de los beneficiarios en el corto y en el medio plazo. Para realizar dicho ejercicio de evaluación usamos las técnicas econométricas de modelos de emparejamiento (*propensity score matching,* PSM) descrito previamente[25].

6.1. Función de emparejamiento

La técnica de evaluación que usamos es un algoritmo conformado por diferentes pasos. En la primera etapa estimamos la probabilidad de ser beneficiario del programa, a partir de la muestra de observaciones compuesta por los individuos beneficiarios del programa –tratados–, y los demandantes de empleo no beneficiarios del programa –controles–. En esta fase, el aspecto más delicado es seleccionar adecuadamente las variables explicativas que nos ayudan a predecir dicha probabilidad. Estas variables deben estar correlacionadas con la probabilidad de ser tratado, pero no pueden ser predictores perfectos de la asignación al tratamiento. Si ponemos este requisito en el contexto del tipo de programa que estamos evaluando, esto implica que es fundamental que las variables seleccionadas identifiquen correctamente «la dificultad de inserción laboral». Otro requisito importante es que las variables que formen parte de la función de emparejamiento no estén afectadas por el programa.

Siguiendo estas pautas básicas, así como usando la información relativa a los criterios de selección del programa, las variables explicativas –descritas la mayoría en el apartado de sección estadística– que forman parte del modelo probabilístico que define la función de emparejamiento son: edad, mujer, nivel educativo (estudios primarios, estudios secundarios, estudios superiores), inmigrante, experiencia laboral

[25] Para ese fin se hace uso del comando psmatch2 disponible en el *software* econométrico STATA (Leuven & Sianesi, 2003).

previa (meses), duración del paro (1-3 meses, 4-6 meses, 6-12 meses, 12-23 meses, >24 meses), beneficiario IMS, tamaño del municipio de residencia y provincia de residencia. La edad del individuo se modeliza a partir de un conjunto de variables ficticias que definen diferentes grupos de edad (<20, 20-25, 25-29, 30-39, 40-49, 50-59). La experiencia laboral previa también se discretiza en las siguientes categorías: sin experiencia laboral, entre 1-12 meses, 13-36 meses, 36-60 meses, >60 meses. El tamaño de municipio de residencia se mide a partir del número de habitantes y recoge las diferentes categorías: <5, 5-20, 20-50, 50-100, >100 mil habitantes.

Para estimar la función de emparejamiento definimos una variable ficticia que toma valor la unidad si el individuo participa en el programa (D=1), y cero si el individuo no participa en el programa (D=0). La función de emparejamiento se estima con un modelo de elección discreta, suponiendo que los errores siguen una distribución valor extremo tipo I, usando como variables explicativas las arriba indicadas. Una vez estimado el modelo, se computan las probabilidades de ser beneficiario del programa para cada individuo, tratado y control. Estas probabilidades se usan para encontrar un grupo de individuos no beneficiarios del programa que sirvan como contrafactual de cada uno de los individuos que sí se han beneficiado del programa. En la construcción de la muestra de contrafactuales usamos el criterio del soporte común. Esta condición garantiza que las observaciones del grupo de tratamiento tengan siempre observaciones de comparación «cercanas» en base a las características socioeconómicas usadas para definir la función de emparejamiento.

Los resultados para construir la función de emparejamiento se presentan en el Cuadro 7. La estimación del modelo econométrico relativo a la probabilidad de asignación al tratamiento permite identificar los factores que más peso tienen en la probabilidad de ser tratado. Estos factores son –en cuanto que influyen positivamente en dicha probabilidad de asignación– ser parado de larga duración, ser perceptor del IMS y residir en municipios pequeños frente a grandes. La estimación de este modelo nos permite también evaluar la calidad del emparejamiento y concluir que la calidad es buena[26].

Tras el cálculo de la probabilidad de tratamiento *(propensity score)*, se realiza el emparejamiento entre las observaciones del grupo de tratamiento y el grupo de control. La ventaja de la amplia muestra de controles se manifiesta en el alto porcentaje de observaciones dentro del soporte común. En el programa Emple@30+, el soporte común es del 95 % (12 tratados y 5.670 controles fuera del soporte común), mientras que en el de Emple@Joven, el soporte común alcanza el 99 % (1 tratado y

[26] El ajuste de los modelos de regresión es bueno (pseudo-R2 Nagelkerke en torno a 0,25). Para el programa 30+, el porcentaje de casos correctos pronosticados como de «tratamiento» se sitúa en el 66 % y en el 90 % para el grupo de control. El porcentaje de «falsos positivos» es del 33 %: casos de «control» pronosticados como de tratamiento. Para el programa joven, el porcentaje de casos correctos pronosticados como de «tratamiento» se sitúa en el 61 % y para el grupo de control se sitúa en el 77 %. El porcentaje de «falsos positivos» es del 38 %: casos de «control» pronosticados como de tratamiento.

CUADRO 7

RESULTADOS DE LA ESTIMACIÓN DE LA FUNCIÓN DE EMPAREJAMIENTO POR TIPO DE PROGRAMA (Modelo Logit: Y = 1 si el individuo es tratado; Y = 0 si el individuo no es tratado)

	30+	Joven
Inmigrante	–0,0647***	–0,176***
Beneficiario IMS	0,238***	0,379***
Duración del paro		
3-6 meses	0,0210***	0,114***
6-12 meses	0,0135***	0,128***
12-23 meses	0,0775***	0,271***
> 24 meses	0,107***	0,331***
Experiencia laboral		
Sin experiencia	–0,0722***	–0,0326***
1-6 meses	–0,00924***	0,0983***
6-12 meses	0,00209	0,0837***
Minusválido	0,00898**	0,0465***
Edad		
40-49	0,00652***	
50-55	0,0176***	
>55	–0,0127***	
>20		0,127***
20-24		0,117***
Educación		
Primaria	–0,0560***	–0,137***
Secundaria	0,0832***	0,203***
Mujer		
Mujer	–0,0580***	–0,0873***
*Mujer * Educación*		
Primaria	0,000830	0,0272**
Secundaria	0,0394***	0,0295***
*Educación * Paro*		
Primaria paro 12-24 meses	0,0509***	0,0733***
Primaria paro >24 meses	0,00932	0,0766***
Secundaria paro 12-24 meses	0,0176**	0,0573***
Secundaria paro >24 meses	0,0252***	0,0417***
*Mujer * Paro*		
Mujer paro 12-24 meses	–0,0168***	0,0572***
Mujer paro >24 meses	–0,00259	0,0551***
Tamaño municipio		
<5 miles	0,000830	0,0272**
5-20 miles	0,0394***	0,0295***
20-50 miles	0,0509***	0,0733***
50-100 miles	0,00932	0,0766***

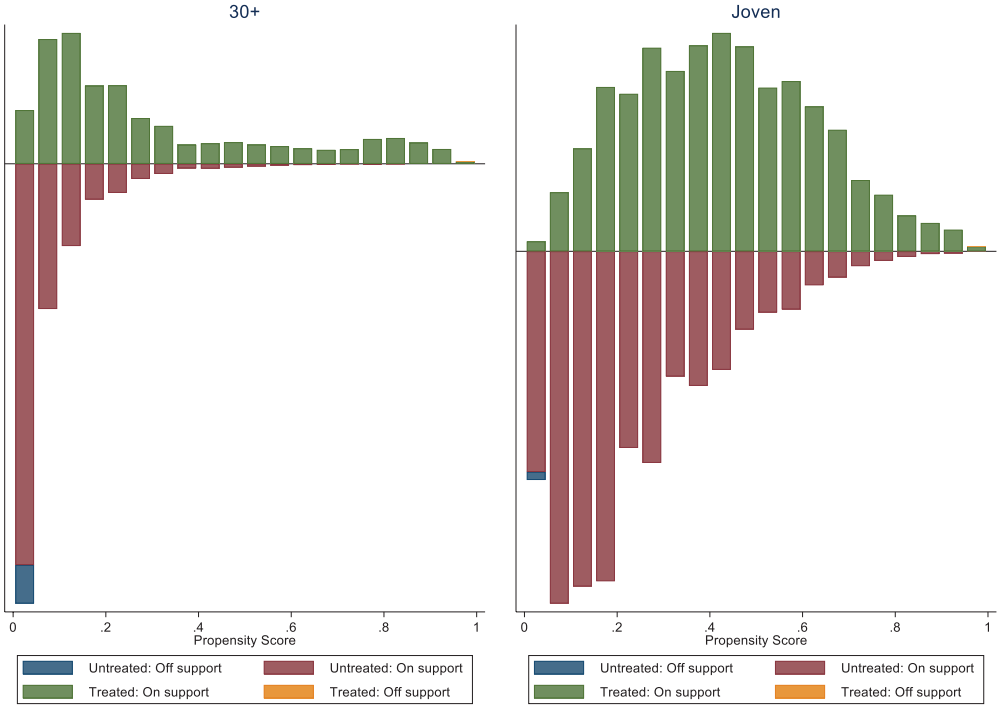
FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.
NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %.

129 controles fuera del soporte común). Esta ventaja también se manifiesta en que el número medio de emparejamiento para cada participante es alto (38), siendo el número mínimo de 3 en el caso del programa Emple@Joven y de 8 en el programa Emple@30+. Los emparejamientos se han realizado con al menos 20 unidades de comparación y solo con aquellas observaciones que estén en un radio de 0,01, para que dichos emparejamientos sean lo más semejantes posibles.

Con el objeto de asegurar la calidad adecuada de la función de emparejamiento, se ha verificado la condición de balanceo (*balancing tests)* en las características observadas entre la muestra de tratados y sus contrafactuales correspondientes. Es decir, que no hay diferencias estadísticamente significativas en la media de cada uno de los regresores entre los tratados y sus contrafactuales, una vez realizado el emparejamiento. Las Figuras 1 y 2 muestran que ambas propiedades se cumplen.

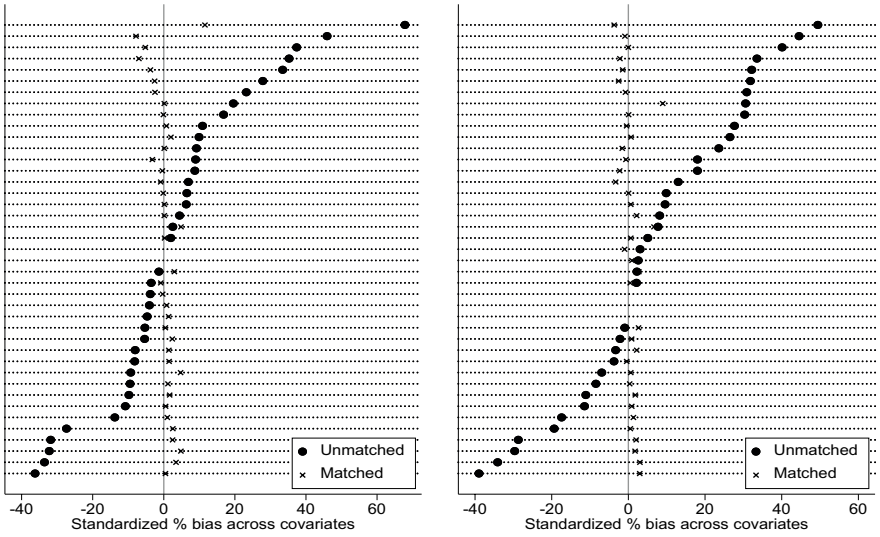
FIGURA 1

EVALUACIÓN DE LA CALIDAD DEL EMPAREJAMIENTO: DISTRIBUCIÓN DE LA PROBABILIDAD DE RECIBIR EL TRATAMIENTO PARA EL GRUPO DE TRATAMIENTO Y EL GRUPO DE CONTROL POR TIPO DE PROGRAMA



FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.

FIGURA 2
EVALUACIÓN DE LA CALIDAD DEL EMPAREJAMIENTO



FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.

6.2. Resultados de la evaluación: efecto medio del tratamiento

Con los grupos de tratados y sus correspondientes contrafactuales, se realiza la evaluación de impacto del programa. Es decir, se comparan los resultados de las variables de interés entre el grupo de tratamiento y los contrafactuales seleccionados. Se seleccionan dos tipos de indicadores de resultados: la empleabilidad de los individuos y la calidad del nuevo empleo. La empleabilidad se mide a partir de la probabilidad de estar empleado seis meses después de ser beneficiario del programa de empleo (horizonte temporal de medio plazo). La calidad del nuevo empleo se mide a partir del tipo de contrato, la duración del contrato y el tipo de ocupación. El tipo de contrato se define en función de que se trate de un contrato temporal o indefinido. La idea es que, en media, un contrato indefinido garantiza una mayor estabilidad laboral que un contrato temporal. El tipo de ocupación se mide a partir de los grupos de cotización del nuevo contrato. Y finalmente, la duración del contrato se define como una variable binaria en base a una duración de menos o más de seis meses (corta duración frente a media/larga duración).

En el Cuadro 8 se presentan los resultados principales del ejercicio de evaluación para la muestra total de individuos y para la submuestra de individuos vulnerables. Más concretamente, se presenta la probabilidad –estimada– de estar empleado seis meses después de haber finalizado el contrato subvencionado para los tratados y sus contrafactuales, el efecto medio del tratamiento sobre los tratados estimado (ATT)

CUADRO 8

EFECTO CAUSAL DEL PROGRAMA SOBRE LA PROBABILIDAD DE OBSERVAR AL INDIVIDUO EMPLEADO SEIS MESES DESPUÉS DEL TRATAMIENTO (SALIDA DEL PARO AL EMPLEO)

		Probabilidad de salir del paro al empleo		Diff	
PSM		Tratado	Control	ATT	T-stat
30+	Todos	0,281	0,293	–0,012***	–2,3
	CIDL	0,211	0,199	0,012***	1,9
Joven	Todos	0,325	0,378	–0,052***	–8,5
	CIDL	0,251	0,263	–0,012**	–1,6

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.
NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %. Boostrap standard errors.

y la significatividad estadística de dicho estimador. Estas estimaciones se presentan para cada programa (30+ y joven).

Lo primero que destaca en estos resultados es que el estimador ATT no parece ser positivo, es decir, no parece que el programa, en media, haya favorecido la salida del paro al empleo para los beneficiarios del programa. Así, para los jóvenes, los efectos estimados son de –5,2 pp para la muestra total y –1,2 pp para la submuestra de individuos con dificultades de inserción laboral. Para la muestra 30+, la probabilidad estimada de estar empleado tras el tratamiento es del 28,1 % para los tratados y, según el modelo, hubiese sido de un 29,3 % para esa muestra de tratados en ausencia del programa. Esto es, el efecto causal estimado es que se reduce la probabilidad de estar empleado en 1,2 pp. Por el contrario, cuando nos centramos en la submuestra de individuos mayores de 30 con dificultades de inserción laboral, la probabilidad de estar empleado es del 21,1 % y hubiese sido del 19,9 % en ausencia del tratamiento. Por tanto, es solo en este caso, parados mayores de 30 años con dificultades de inserción laboral, que encontramos un efecto causal positivo y significativo de en torno a 1,2 pp.

Los resultados indican, por tanto, que la política no parece estar beneficiando la empleabilidad de sus beneficiarios salvo para colectivos específicos. Los resultados presentados están en línea con la mayoría de la literatura empírica resumida a partir de los metaanálisis de Card, Kluve y Weber (2010, 2018) y Kluve (2010). En ellos se concluye que, en media, las políticas activas consistentes en la creación de empleo público no suelen llevar beneficios asociados a mayores niveles de empleabilidad de sus participantes. El análisis presentado no permite ofrecer una conclusión clara sobre los motivos o mecanismos que expliquen la falta de eficacia de esta política. La literatura empírica hace referencia a dos tipos de mecanismos, uno básicamente por el lado de la oferta de trabajo y otro por el lado de la demanda. En el primer caso,

la literatura habla de los efectos de bloqueo *(lock in effects)*. Esto es, los individuos quedan atrapados en este tipo de programas cuando la duración del programa es larga. Por otro lado, también se habla de los efectos estigma *(stigma effects)* que pueden sufrir los participantes de este tipo de políticas (Caliendo, 2006).

No obstante, es importante profundizar en el análisis, dado que dentro del colectivo de beneficiarios hay individuos con diferentes características y niveles potenciales de exclusión social, y esto puede dar lugar a una gran heterogeneidad de efectos causales. Por ello, en el Cuadro 9 se presentan los efectos medios –causales– del programa para diferentes tipos de individuos según género, nivel de estudios, edad, duración del episodio de paro, nacionalidad, beneficiario del IMS, por experiencia laboral y beneficiario del anterior programa de iniciativa de empleo local. En general, son todas ellas características socioeconómicas que nos ayudan a identificar mejor las situaciones de vulnerabilidad social y laboral.

Los resultados presentados en el Cuadro 9 apuntan la existencia de importantes diferencias en la incidencia del programa sobre la empleabilidad de los individuos beneficiarios del mismo. Comenzando con la duración del paro, vemos que los efectos positivos del programa aparecen para los individuos mayores de 30 años con episodios de paro de más de dos años. Para este colectivo, el programa ha aumentado la empleabilidad a medio plazo en 2,1 pp. Para otros colectivos, tales como los individuos mayores de 30 años sin experiencia laboral, los resultados también parecen ser positivos al aumentarles su empleabilidad en 3,1 pp. Por el contrario, para el resto de los colectivos y, en general, para los menores de 30 años, el programa no parece haber favorecido los niveles de empleabilidad de los beneficiarios.

Los resultados del Cuadro 9 también nos permiten indicar algunos factores que pueden ayudarnos a entender parte de los resultados encontrados. En concreto, es interesante destacar los impactos fuertemente negativos encontrados para los beneficiarios del IMS o para los individuos que también se habían beneficiado de iniciativas locales de empleo de convocatorias anteriores (programa en 2014, Tratado14=1). Ambos resultados apuntan a la existencia de que una parte de los beneficiarios se encuentran fuertemente vinculados a estos programas de ayuda social *–lock in effects–*, lo que puede desfavorecer una búsqueda activa de empleo y una menor tasa de llegada de ofertas.

Una vez presentados los resultados relativos a los efectos del programa sobre la empleabilidad de los individuos beneficiarios del mismo, pasamos a evaluar sus efectos a medio plazo y largo plazo. Más concretamente, se evalúa si el programa ha podido ayudar a mejorar la calidad del nuevo empleo (frente al empleo que hubiesen conseguido en ausencia del programa). Como se indicó anteriormente, las variables que usamos para medir la calidad del nuevo empleo son: tipo y duración del nuevo contrato, así como la cualificación del empleo.

El ejercicio de evaluación sigue los mismos pasos que los seguidos para evaluar la probabilidad de salir del paro. Los principales resultados se presentan en el Cuadro 10. La principal idea que emerge de dicho cuadro es que, en media, las características del empleo de los tratados pueden ser algo mejores que la de los controles.

CUADRO 9

EFECTO CAUSAL DEL PROGRAMA SOBRE LA PROBABILIDAD DE OBSERVAR AL INDIVIDUO EMPLEADO SEIS MESES DESPUÉS DEL TRATAMIENTO POR CARACTERÍSTICAS SOCIOECONÓMICAS

		30+	Joven
		ATT	ATT
Género	Mujer	–0,003	–0,062***
	Hombre	–0,020***	–0,027***
Formación	Estudios primarios	–0,002	–0,054**
	Estudios secundarios	–0,008	–0,047***
	Estudios superiores	–0,023***	–0,035***
Edad	< 20 años	–	–0,003
	20-24 años	–	–0,040***
	25-29 años	–	–0,057***
	30-39 años	0,002	–
	40-49 años	–0,004	–
	50-59 años	–0,033**	–
Submuestra CDIL			
Duración del paro	< 6 meses	–0,057***	–0,086***
	6-12 meses	–0,045***	–0,059***
	12- 24 meses	–0,007	–0,024***
	> 24 meses	0,021***	–0,010
Inmigrante		0,006	–0,016
Nativo		–0,015***	–0,045***
Beneficiado IMS	Sí	–0,028***	–0,096***
	No	–0,008***	–0,047***
Minusválido		0,010	–0,006
No minusválido		–0,014***	–0,047***
Experiencia laboral	Sin experiencia.	0,031***	0,008
	Con experiencia	–0,021***	–0,079***
Programa 2014	Tratado14=1	–0,086***	–0,096***
	Tratado14=0	–0,010**	–0,035***

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.
NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %. Boostrap standard errors.

Así, el Cuadro 10 muestra que el programa favorece la empleabilidad en empleos con contratos a tiempo completo y reduce la probabilidad de que dichos contratos sean de corta duración y/o de baja cualificación. Estos resultados podrían estar indicando que, para los colectivos donde el programa parece tener un impacto positivo en la empleabilidad, la inserción en el mercado de trabajo podría estar resultando a su vez en empleos de mayor estabilidad y duración. Sin embargo, hay que tomar este resultado con precaución dado que hemos visto como son pocos los colectivos que se benefician de estas medidas en términos de mayor empleabilidad, cuando lo comparamos con individuos comparables y no beneficiarios de la política.

En conclusión, por tanto, no se observa que, en media, el programa haya causado un aumento en la probabilidad de salir del paro al empleo seis meses después de haber participado en el programa. La heterogeneidad de los resultados es importante para el colectivo de mayores de 30 años, sobre todo, ya que hay determinados colectivos para los que los efectos sí parecen haber sido favorables, como son los parados de muy larga duración y las mujeres con estudios primarios/secundarios.

Por último, es importante tener en cuenta que, con el presente ejercicio de evaluación, no se evalúa si el programa ha evitado que los beneficiarios abandonen el mercado de trabajo, es decir, los efectos sobre la tasa de actividad. Dado que parte de los colectivos beneficiarios son individuos con altas dificultades de inserción laboral, un indicador complementario de esta evaluación sería evaluar sus efectos sobre la probabilidad de abandonar el mercado de trabajo. La falta de información estadística para medir esta variable de resultado impide arrojar luz a este respecto.

CUADRO 10

EFECTO CAUSAL SOBRE OTROS INDICADORES DE EMPLEABILIDAD, SUBMUESTRA DE REEMPLEADOS

		Todos ATT		**CDIL ATT**		**Paro > 24 meses ATT**	
		30+	**Joven**	**30+**	**Joven**	**30+**	**Joven**
Tiempo completo		0,072***	0,058***	0,052***	0,051***	0,034***	0,002
Duración del contrato	< 3 meses	–0,071***	–0,065***	–0,059***	–0,076***	–0,057***	–0,079***
	3-6 meses	–0,040***	–0,016***	–0,014***	–0,014***	–0,031***	–0,020
	Fecha final	–0,138***	–0,090***	–0,163***	–0,096***	–0,143***	–0,082***
Cualif. del empleo	Alta	0,008	0,004	0,022***	0,016*	0,006	0,048***
	Media	–0,027***	–0,015**	0,051***	0,010	0,076***	–0,023
	Baja	0,019**	0,011	–0,072***	–0,026**	–0,082***	–0,026

FUENTE: Elaboración propia a partir de los datos proporcionados por el SAE.
NOTA: *** Significativo al 1 %. ** Significativo al 5 %. * Significativo al 10 %. Boostrap standard errors.

7. Conclusiones

La evaluación de las PAE viene justificada por la necesidad de determinar la efectividad y la eficacia de estos programas. La efectividad en el logro de los objetivos propuestos y la eficacia en términos de relación entre el logro de objetivos y recursos invertidos para alcanzarlos. Además, solo a través de un análisis detallado de los programas se determinará sobre qué poblaciones tendrán una mayor incidencia y cuáles van a ser los mejores procesos para llevarlos a cabo, para su correcta puesta en marcha y ejecución.

La revisión de la literatura ha puesto de manifiesto que las iniciativas locales de empleo no funcionan en todas las circunstancias y con todos los desempleados. El diseño, la cobertura y los objetivos de estos programas, y el modo en que se aplican, pueden afectar mucho en términos de eficacia y eficiencia. No existe una única política que sirva como herramienta universal para mejorar las perspectivas laborales de los desempleados.

Las iniciativas locales de empleo analizadas en este trabajo se han diseñado de cara a lograr un alto grado de cobertura, tanto por la participación de las entidades beneficiarias (ayuntamientos andaluces) como por la participación de personas desempleadas no ocupadas inscritas como demandantes de empleo. Una gran parte de estos últimos se encuentran en situación de especial vulnerabilidad social, por lo que podemos decir que se trata de programas con un alcance de gran valor social. Desde un punto de vista de la eficacia de la iniciativa, se podría afirmar que ha logrado algunos de los objetivos que se planteaban como es fomentar la contratación en el ámbito local y dinamizar el mercado de trabajo durante la duración de los proyectos. No obstante, no parece tan clara la contribución del mismo respecto a la mejora de la empleabilidad futura de las personas participantes una vez finalizado el contrato estipulado, al menos no en el caso de los jóvenes. Es decir, parece que estos programas no son efectivos para el logro de la reinserción de una gran parte de los colectivos objetivo de dichas políticas. Sin embargo, sí que se observan algunos resultados positivos para desempleados mayores de 30 años con especiales dificultades de inserción. Por tanto, nuestros resultados están apuntando a la necesidad de ajustar mucho más los objetivos y, sobre todo, los colectivos destinatarios de estas medidas, para conseguir un impacto mayor sobre la empleabilidad y la inserción de los mismos.

A pesar de que los resultados presentados no permiten hablar de una ganancia clara en la empleabilidad de los colectivos tratados, hay que precisar que, para valorar de manera completa estas iniciativas, sería necesario tener en cuenta también sus efectos sobre la tasa de salida del paro a la inactividad, de cara a evaluar si logra reducir la subida de la tasa de inactividad por el efecto «trabajador desanimado». En términos generales, las PAE son muy necesarias para combatir el desempleo, que de forma tan grave afecta al mercado de trabajo español. Esto es especialmente importante para quienes están en los extremos de su carrera laboral (jóvenes y mayores de 45 años) y para los colectivos más vulnerables (bajo nivel de estudios, desempleados de larga duración, etc.). Por tanto, los ejercicios de evaluación de las PAE como el

presentado en este artículo, deben considerarse como un instrumento de apoyo al diseño de dichas políticas para futuras convocatorias.

Finalmente, indicar que las evaluaciones cuasiexperimentales como las presentadas en este artículo son todavía escasas en España. La evaluación presentada es relativamente sencilla y de coste reducido, pues los datos ya existen y lo único que se necesita es una extracción correcta de las bases de datos disponibles y un análisis que, hoy en día, es mucho más sencillo de acometer gracias a diferentes paquetes estadísticos.

Referencias bibliográficas

Arellano, F. A. (2010). Do training programmes get the unemployed back to work? A look at the Spanish experience. *Revista de Economía Aplicada*, *18*(53), 39-65. https://www.researchgate.net/publication/47647401_Do_Training_Programmes_Get_the Unemployed Back to Work A Look at the Spanish Experience

Arranz, J. M., García Serrano, C., & Hernanz, V. (2013). Active labour market policies in Spain: A macroeconomic evaluation. *International Labour Review*, *152*(2), 327-348. https://doi.org/10.1111/j.1564-913X.2013.00182.x

Blázquez, M., Herrarte, A., & Sáez, F. (2019). Training and job search assistance programmes in Spain: The case of long-term unemployed. *Journal of Policy Modeling*, *41*(2), 316-335. https://doi.org/10.1016/j.jpolmod.2019.03.004

Caliendo, M. (2006). *Microeconometric evaluation of labour market policies*. Springer Science & Business Media.

Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *The economic journal*, *120*(548), F452-F477. https://doi.org/10.1111/j.1468-0297.2010.02387.x

Card, D., Kluve, J., & Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894-931. https://doi.org/10.1093/jeea/jvx028

Cueto, B., & Mato, F. J. (2009). A nonexperimental evaluation of training programmes: regional evidence for Spain. *The Annals of Regional Science*, *43*(2), 415-433. https://doi.org/10.1007/s00168-008-0214-2

De la Rica, S. (2015). Políticas activas de empleo: una panorámica. *Fedea policy papers*, *1*, 1-32. https://documentos.fedea.net/pubs/fpp/2015/01/FPP2015-01.pdf

Escudero, V., Kluve, J., López Mourelo, E., & Pignatti, C. (2019). Active labour market programmes in Latin America and the Caribbean: Evidence from a meta-analysis. *The Journal of Development Studies*, *55*(12), 2644-2661. https://doi.org/10.1080/00220388.2018.1546843

Felgueroso, F., García-Pérez, J. I., & Jiménez, S. (2018). *Perfilado estadístico: un método para diseñar políticas activas de empleo*. Editorial Centro de Estudios Ramón Areces. https://www.fundacionareces.es/recursos/doc/portal/2018/05/03/perfilado-estadistico.pdf

García-Pérez, J. I., & Rebollo-Sanz, Y. F. (2009a). The use of Permanent Contracts Across Spanish Regions: Do regional wage subsidies work? *Investigaciones Económicas, 33*(1), 97-130. https://www.redalyc.org/pdf/173/17312887004.pdf

García-Pérez, J. I., & Rebollo-Sanz, Y. F. (2009b). Do wage subsidies affect the subsequent employment stability of permanent workers?: the case of Spain. *Moneda y Crédito,* 228, 65-103. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.529.8719&rep=rep1&type=pdf

Gerfin, M., & Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. *The Economic Journal*, *112*(482), 854-893. https://doi.org/10.1111/1468-0297.00072

Hohmeyer, K., & Wolff, J. (2010). Direct job creation revisited. Is it effective for welfare recipients and does it matter whether participants receive a wage? *IAB*, Discussion Paper No. 21/2010, 1-62. http://doku.iab.de/discussionpapers/2010/dp2110.pdf

INAEM, Instituto Aragonés de Empleo. (2015). *Estudio sobre la valoración económica y social de las políticas activas del Instituto Aragonés de Empleo*. https://inaem.aragon.es/sites/default/files/inaem_2010-2014_25112015.pdf

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour economics*, *17*(6), 904-918. https://doi.org/10.1016/j.labeco.2010.02.004

Kuddo, A. (2009). Employment services and active labor market programs in Eastern European and Central Asian countries. *SP, Social Protection* Discussion Paper No. 0918, 1-87. http://documents.worldbank.org/curated/en/801251468170040578/Employment-services-and-active-labor-market-programs-in-Eastern-European-and-Central-Asian countries

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of econometrics*, *125*(1-2), 305-353. https://doi.org/10.1016/j.jeconom.2004.04.011

Ramos, R., Suriñach, J., & Artís, M. (2009). *La efectividad de las políticas activas de mercado de trabajo para luchar contra el paro. La experiencia de Cataluña* (Working Papers No. 2009/19). Research Institute of Applied Economics. http://www.ub.edu/irea/working_papers/2009/200919.pdf

Rebollo-Sanz, Y. F. (2018). El Modelo de Perfilado Estadístico: una herramienta eficiente para caracterizar a los demandantes de empleo. En F. Felgueroso, J. I. García-Pérez, & S. Jiménez (Eds.), *Perfilado estadístico: un método para diseñar políticas activas de empleo* (pp. 17-93). Editorial Centro de Estudios Ramón Areces. https://www.fundacionareces.es/recursos/doc/portal/2018/05/03/perfilado-estadistico.pdf

Roy, A. S., & Wong, G. (2000). Direct job creation programs: Evaluation lessons on cost-effectiveness. *Canadian Public Policy/Analyse de Politiques*, *26*(2), 157-169. https://doi.org/10.2307/3552553

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

El perfilado estadístico como instrumento para la evaluación del impacto del programa Incorpora*

Statistical Profiling: an instrument for evaluating the impact of Incorpora

Jorge Casanova
Compass Lexecon y FEDEA

Florentino Felgueroso
Universidad de Oviedo y FEDEA

José Ignacio García Pérez
Universidad Pablo de Olavide y FEDEA

Sergi Jiménez-Martín
Universitat Pompeu Fabra y FEDEA

Resumen

En este artículo presentamos un ejercicio de evaluación causal en el contexto del perfilado estadístico aplicado a un programa de empleo creado por la Fundación "La Caixa" y denominado Incorpora. Este programa consiste en un mecanismo de apoyo y recolocación en el mercado de trabajo de personas en riesgo de exclusión. Nuestros resultados apuntan a que la formación impartida dentro del programa contribuye a una mejora general de la empleabilidad media de los participantes en el programa, pero se observa cómo este efecto decae en el tiempo que el beneficiario está desempleado. Por otra parte, encontramos que el programa Incorpora mejora la probabilidad de encontrar y retener el empleo, así como el tiempo trabajado e incluso el salario una vez reempleado, para todos los colectivos analizados sin importar el género, nacionalidad, sector previo de empleo o experiencia laboral previa. Concretamente, la empleabilidad aumenta entre 3 y 4 puntos porcentuales gracias al programa y el efecto sobre el número de meses acumulados en el empleo se sitúa, dependiendo de la cohorte, entre 3 y 4 meses tres años después de salir del mismo.

Palabras clave: *perfilado estadístico, análisis causal, contrafactual, tratados y grupo de control.*
Clasificación JEL: *C35, H43, J15, J64.*

Abstract

In this article, we present a causal evaluation exercise in the context of statistical profiling applied to an employment program created by La Caixa Foundation and called Incorpora. This program consists of a mechanism of support and repositioning in the labor market of people at risk of exclusion. Our results suggest that the training given within the program contributes to a general improvement in the average employability of the participants in the program, but it is observed that this effect declines with unemployment duration. On the other hand, we find that the Incorpora program improves the probability of finding and retaining employment as well as the time worked and even the salary once re-employed for all the groups analyzed regardless of gender, nationality, previous sector of employment or previous work experience. Specifically, employability increases between 3 and 4 percentage points thanks to the program and the effect on the number of months accumulated in employment is situated, depending on the cohort, between 3 and 4 months three years after leaving it.

Keywords: *statistical profiling, causal evaluation, counterfactual, treated and control groups.*

* Agradecemos a la Fundación "La Caixa" la financiación recibida para elaborar este estudio y el permiso para publicar algunos resultados anticipándonos a la publicación del informe final. Las opiniones contenidas en este documento son responsabilidad exclusiva de los autores y no pueden atribuirse a Compass Lexecon ni a cualquier otra parte.

DOI: https://doi.org/10.32796/cice.2021.102.7316
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

1. Introducción

La situación de alto desempleo y la persistencia en dicho estado que sufren muchos de los trabajadores que son despedidos o que buscan empleo por primera vez hacen cada vez más necesario la inversión en políticas de formación para hacer frente al desajuste entre las capacidades y competencias de una mano de obra en situación de desempleo que no corresponden con las necesidades y demandas de la economía productiva. Asimismo, una estancia prolongada en el desempleo puede deteriorar no solo las competencias de los desempleados, sino la motivación y actitud de estos hacia la búsqueda de empleo. Por ello, también se hace necesario reforzar la motivación para la búsqueda activa, el desarrollo de estrategias de activación y el fomento de iniciativas de emprendimiento empresarial o de incentivos a la contratación. Todas estas políticas o servicios se desarrollan normalmente por las oficinas públicas de empleo en base a lo que se conoce como políticas activas de empleo. También hay algunas instituciones privadas, normalmente fundaciones u organizaciones sin ánimo de lucro, que dedican parte de sus esfuerzos a desarrollar programas de apoyo a desempleados que pueden ser enmarcados dentro de este tipo de políticas activas de empleo. Así, en el conjunto de España, según datos de la OCDE, se dedica a estas políticas en torno al 0,7 % del PIB (datos de 2018), cifra superior en dos décimas a la media del conjunto de países de dicha organización, pero lejana de lo que dedican países como Suecia (1,1 %) o Dinamarca (1,81 %). Los altos niveles de desempleo en España y, sobre todo, la persistente duración de este para ciertos colectivos, hacen muy necesario dedicar un mayor esfuerzo a identificar cuál es el efecto esperado de cada una de estas políticas en los distintos colectivos que componen el conjunto de demandantes de empleo. No es solamente por razones de eficiencia económica que se hace necesario analizar en qué partidas es más eficiente focalizar dicha inversión, sino que también se hace necesario por razones de oportunidad. Una vez clasificados los demandantes en función de su nivel esperado de empleabilidad, sería muy conveniente analizar qué políticas o servicios de entre los disponibles en una oficina de empleo son más eficaces de cara a reinsertar al demandante de nuevo en el *pool* de empleados. Por ello, necesitamos conocer qué programas o servicios son más eficaces en base a los efectos que cada uno de ellos tiene sobre el desempeño de sus usuarios en el mercado de trabajo durante y después de recibir dicho servicio.

Este análisis de desempeño se realiza en la literatura económica mediante técnicas de evaluación causal. Estas técnicas se han desarrollado dentro de una literatura de Evaluación de Políticas Públicas, en creciente expansión en los últimos años. Esta área de investigación persigue la obtención de información útil para valorar la racionalidad, coherencia, eficacia, idoneidad e impacto de las intervenciones públicas. Como herramienta al servicio del gestor público, la utilización de técnicas de evaluación mejora la intervención pública y aumenta su utilidad social; no solo por la información que suministra a todos aquellos interesados en conocer los resultados de las políticas evaluadas, sino por la posibilidad de aplicar sus conclusiones para retroalimentar y mejorar el diseño y elaboración de nuevas intervenciones, y así mejorar los resultados finales y sus impactos.

Esta metodología se complementa muy bien con las técnicas de perfilado estadístico que cada vez se usan más para clasificar a los demandantes de empleo en función de su empleabilidad (Desiere *et al*., 2019). Concretamente, tras clasificar a estos demandantes en distintos segmentos en función de su empleabilidad, se distingue dentro de cada segmento entre colectivos elegibles y no elegibles de cada uno de los servicios a analizar y se analiza, mediante técnicas de emparejamiento o de análisis de discontinuidad, si se observan efectos significativos en cada segmento de las políticas o servicios a evaluar. El objetivo es comparar individuos elegibles con los no elegibles, pero de un mismo nivel de empleabilidad. Esto quiere decir que se estima el impacto comparando aquellos que reciben el servicio con aquellos que no, pero tenían las mismas probabilidades de encontrar un empleo en ese momento.

En este documento se presenta un ejercicio de evaluación causal en el contexto del perfilado estadístico aplicado a un programa de empleo creado por la Fundación «La Caixa» denominado Incorpora. Este programa consiste en un mecanismo de apoyo y recolocación en el mercado laboral de personas en riesgo de exclusión. A través de un asesoramiento individualizado y con recursos de formación y búsqueda de empleo, la intención del programa es lograr la inserción laboral y el mantenimiento del empleo de desempleados de difícil inserción, así como de colectivos en riesgo de exclusión social.

Concretamente, en nuestro análisis del programa Incorpora construimos una metodología de perfilado estadístico que podrá ayudar a incrementar la efectividad de la intermediación hecha por los propios técnicos del programa. A partir de los datos suministrados por la plataforma Incorpora, obtenemos un indicador cuantitativo que aproxima el nivel de empleabilidad del beneficiario, esto es, la verosimilitud de una eventual reincorporación al mercado de trabajo en un plazo dado. Una vez creado dicho indicador, evaluamos, entre otros aspectos, el efecto de la formación sobre dicho nivel de empleabilidad. En segundo lugar, y como objetivo central, abordamos el análisis del impacto causal a medio plazo del programa Incorpora sobre diversos indicadores de inclusión laboral de los beneficiarios, en comparación con un amplio conjunto de individuos que no formaron parte de este, pero han tenido vidas laborales previas muy similares y que, por tanto, pueden ser considerados como grupo de control.

Los resultados de la primera parte apuntan a que la formación impartida dentro del programa contribuye a una mejora general de la empleabilidad media de los participantes en el programa, pero se observa cómo este efecto decae en el tiempo que el beneficiario está desempleado. No obstante, observamos que la asignación de los cursos no prioriza sistemáticamente a los beneficiarios más vulnerables dentro de cada colectivo.

Por otra parte, encontramos que el programa Incorpora mejora los parámetros de integración laboral (probabilidad de encontrar y retener el empleo, tiempo trabajado e incluso salarios) a medio plazo de todos los colectivos analizados sin importar el género, nacionalidad, sector previo de empleo o experiencia laboral previa. Asimismo, la permanencia en el empleo es mayor para los beneficiarios del programa que

para los individuos del grupo de comparación o control, generando también mayores ganancias tanto en términos de salario como de meses acumulados en el empleo.

El uso de mecanismos de apoyo y formación para el empleo no es una novedad. En un metaanálisis, Card, Kluve y Weber (2018) evalúan 199 estudios y muestran como las políticas de formación y del sector privado obtienen mejores resultados en varias dimensiones (incluido salarios y tiempo empleado) tanto a medio como largo plazo.

En el caso de España, la evidencia es escasa debido a la falta de datos y la disposición a evaluar políticas públicas. Si bien es cierto que existen programas de inserción laboral, no tenemos conocimiento de algún estudio que los evalúe. Por el contrario, la literatura se ha centrado mayormente en estudiar políticas activas de empleo enfocadas en incentivar el empleo privado y en programas de formación de parados. Sobre este último, podemos mencionar el estudio de Arellano (2010), que evaluó los programas de formación dirigidos a desempleados por el Plan Nacional para la Formación y la Recolocación. Usando metodologías de estimación causal, encuentra que los cursos de formación fueron eficaces para reducir la duración de desempleo posterior entre 24 y 39 días.

La estructura del resto del artículo es la siguiente. Primero, describiremos el programa y su evolución. Segundo, dividiremos el análisis en dos partes (perfilado y análisis causal) y, sobre estas, haremos una descripción de los datos y la metodología utilizada. Terminamos con unas breves conclusiones.

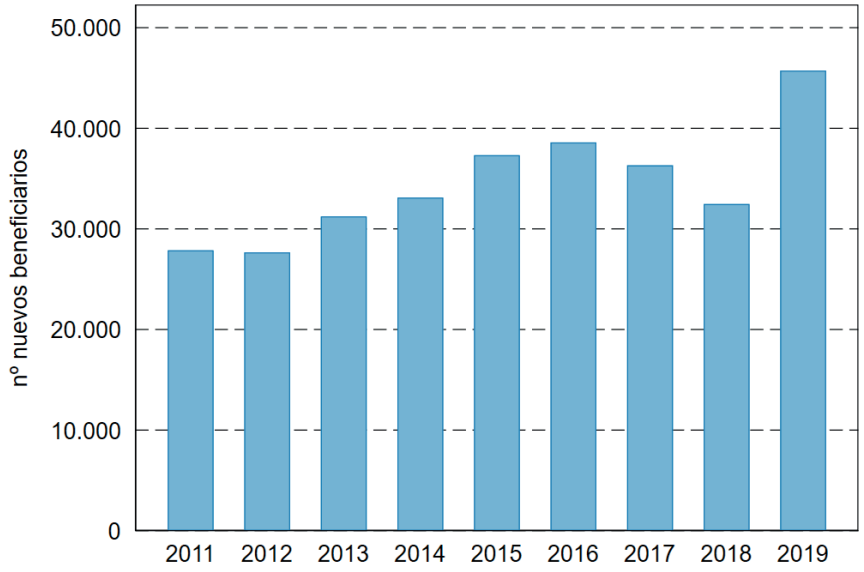
2. Descripción del programa

Desde su fundación en 2006, el programa Incorpora ha apoyado a 330.814 personas en riesgo de exclusión social a través de una red de 617 entidades y 3.307 técnicos especializados[1]. Tal y como muestra la Figura 1, la media de incorporaciones al programa está en torno a las 40.000 anuales, lo que, a su vez, se ha visto complementado con una oferta de más de 450 cursos anuales en promedio que ofrecen herramientas necesarias en la búsqueda del empleo y en todo tipo de ocupaciones. Por otro lado, y gracias a la labor de las entidades asociadas al programa, se realiza una intensa tarea de búsqueda de nuevas oportunidades de acuerdos con empresas de todos los sectores, fruto del cual el número de ofertas de empleo acumuladas ha crecido hasta los 206.636 en 2019, habiéndose firmado hasta esta fecha más de 7.100 convenios con empresas colaboradoras. Esto ha hecho además que el número de inserciones laborales creciera tanto extensivamente como intensivamente, con 14.006 nuevos contratos laborales solo en 2019 y con una línea complementaria de apoyo al autoempleo que ha logrado la creación de microempresas en todo el territorio.

Desde el 2016, Incorpora complementa además el apoyo a los beneficiarios con una nueva línea de autoempleo que asesora a aquellos interesados en emprender.

[1] Datos actualizados en diciembre del 2019.

FIGURA 1
EVOLUCIÓN DEL NÚMERO DE ALTAS DE BENEFICIARIOS



FUENTE: Elaboración propia en base a datos programa Incorpora.

Esto ha resultado en la creación de nuevas empresas en todo el territorio y que, además, se construyen sobre la base de una estrategia planificada para su éxito.

La participación en Incorpora funciona de la siguiente manera: cualquier interesado puede acceder acudiendo a una de las entidades sociales donde, en primer lugar, se tomarán datos para hacer una evaluación integral de su situación. A partir de entonces, Incorpora mantiene el acompañamiento a la persona durante todo el proceso de integración sociolaboral. Permite a cada persona en situación o en riesgo de exclusión social definir un itinerario propio, con el asesoramiento del personal técnico. Incorpora impulsa y da apoyo a la persona, cualquiera que sea su punto de partida, para avanzar hacia la inserción laboral, por cuenta ajena o propia, ya que el autoempleo también está contemplado dentro de las líneas de actuación del programa.

El objetivo del presente trabajo se concreta en evaluar la efectividad de la metodología Incorpora. Para alcanzar nuestros objetivos, dividimos el análisis en dos partes.

En la primera parte, partimos de un modelo de perfilado estadístico basado con los datos de carácter mensual suministrados por la plataforma del programa Incorpora. Los resultados de la estimación son predicciones estadísticas derivadas de un modelo econométrico estimado exclusivamente a partir de la información de los beneficiarios. Todos estos datos son usados para obtener un indicador cuantitativo que aproxime el nivel de empleabilidad del beneficiario, esto es, la distancia con su probable reincorporación al mercado de trabajo.

Una de las dificultades para evaluar el efecto de la formación es la posible endogeneidad en la asignación de esta por parte de los técnicos del programa. Al haber cierta tendencia a formar parte de determinados colectivos, es necesario ajustar las regresiones añadiendo instrumentos que tengan relación con dicha formación, como son, por ejemplo, las experiencias de empleo y de formaciones previas.

En la segunda parte, hacemos un análisis causal con datos trimestrales comparando beneficiarios del programa Incorpora con personas activas de la muestra continua de vidas laborales de características similares y que se pueden considerar grupos de control. Para los colectivos de inmigrantes, discapacitados, jóvenes y personas de 30 a 44 años y mayores de 45, creamos grupos de tratamiento y control a partir de técnicas de análisis causal y, posteriormente, estimamos los efectos del programa en los tres-cinco años posteriores a la entrada en el mismo.

3. Perfilado estadístico

3.1. Datos

Incorpora posee una plataforma *online* a través de la cual todas las entidades registran los seguimientos de beneficiarios, las ofertas de empleo y los cursos de formación. Los datos provienen de dos ficheros distintos: uno referido a las características de los beneficiarios y otro a sus acciones. En el primer fichero, la información es única y se refiere a características demográficas, sociales (colectivos) y de situación previa a la incorporación.

En el segundo fichero, la fecha de cada acción indica el momento en que el técnico realiza el registro en la plataforma y, por tanto, se asume que entre una acción y otra no ha habido más cambios que los que se indican. A su vez, a partir de estos datos se obtiene la situación laboral de cada beneficiario, que será utilizada para generar la variable dependiente en la muestra de estimación y, también, los valores de la mayoría de las variables explicativas del modelo de empleabilidad.

Una de las ventajas de la plataforma es que permite detallar el tipo de acción que tiene el individuo. No obstante, y para simplificar el análisis, en una primera aproximación nos centraremos en agregar las acciones en seis categorías (alta, parado, formación, empleo, baja o censura).

– Alta: momento en que el individuo se registra en el programa (en el fichero se define como fecha de incorporación).
– Parado: momento en que el individuo no se encuentra inscrito en algún curso de formación o está trabajando. Esta categoría incluye acciones de asesoramiento puntual o entrevistas.
– Formación: los beneficiarios que se inscriban en un curso de formación entran dentro de esta categoría hasta que finaliza o cambian de acción en los siguientes registros.

– Empleo: los beneficiarios que estén trabajando (tanto por cuenta ajena como por cuenta propia) o realicen prácticas de empresa, entran en esta categoría.
– Baja o censura: en algunos casos, los técnicos registran la última acción del beneficiario para indicar que este se dio de baja del programa. En caso contrario, se desconocen las acciones a partir de entonces y se categorizan como censura.

Existen consideraciones metodológicas adicionales que surgieron durante el análisis y que exponemos a continuación.

En primer lugar, beneficiarios repetidos donde el identificador es el mismo pero las características son distintas. En este caso, nos quedamos con un solo registro (no hay un patrón que se repita y, por tanto, tampoco lo es la selección). En segundo lugar, individuos menores de 15 años y mayores de 65. Ninguno de estos casos se considera dentro del análisis porque el año de nacimiento es, en su mayoría, un valor atípico. En tercer lugar, múltiples acciones dentro de un mismo mes. Al ser un fichero de información mensual, nos quedamos con el último registro de cada mes. En caso de tener la misma fecha, se priman acciones referidas al empleo o la formación frente al paro.

Finalmente, individuos con un colectivo no asignado, pero características asignables. Hay beneficiarios cuyo colectivo no estaba asignado o, estando dentro de la categoría de «Otros», poseía características similares a otros. Para homogeneizar los grupos, solo se considera dentro de «Otros» a personas entre 31 y 44 años (edad *prime*) y con nacionalidad española (o sin asignar).

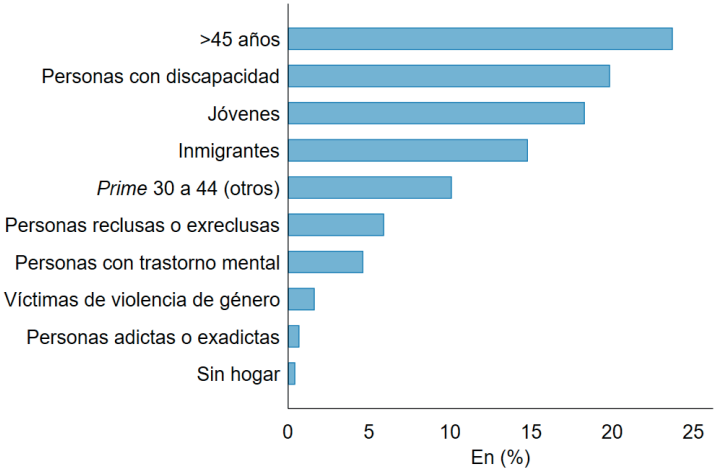
Una vez aplicadas estas restricciones, la muestra de análisis considera como beneficiarios aquellos incorporados al programa en el periodo 2011-2019. Este colectivo incluye un total de 302.632 personas, siendo entre estos los mayores de 45 años y los discapacitados aquellos con mayor peso. Son en su mayoría españoles (véase Figura 2), seguido de latinoamericanos, y mayoritariamente se han dado de alta en los grupos territoriales de Barcelona, Madrid o Andalucía. Por grupos de edad, la mitad de los beneficiarios tienen entre 31 y 50 años y, si lo vemos por colectivo, la distribución está más concentrada entre los 31 y 40 años para colectivos cuya clasificación no depende de la edad.

En cuanto al género, la distribución por colectivos es bastante desigual en casos como las personas adictas o exadictas o los reclusos o exreclusos (véase Figura 2). Esta heterogeneidad es la que buscamos captar al momento de estimar el efecto de la formación ya que nos permite, además, saber a quién beneficia esta, más o menos, y cómo podría enfocarse en el futuro.

Al ser la empleabilidad nuestro objeto de estudio, analizamos también gráficamente cómo se distribuye el porcentaje de beneficiarios que encuentran empleo por las características ya consideradas. En general, como se recoge en la Figura 3, no parece haber diferencias muy significativas entre cualquiera de las categorías, es decir, aproximadamente el 50 % de los beneficiarios consiguieron algún empleo durante su periodo de pertenencia al programa.

FIGURA 2
DESCRIPTIVOS SOBRE LOS BENEFICIARIOS

a) Porcentaje de beneficiarios por colectivo



b) Distribución del género por colectivo

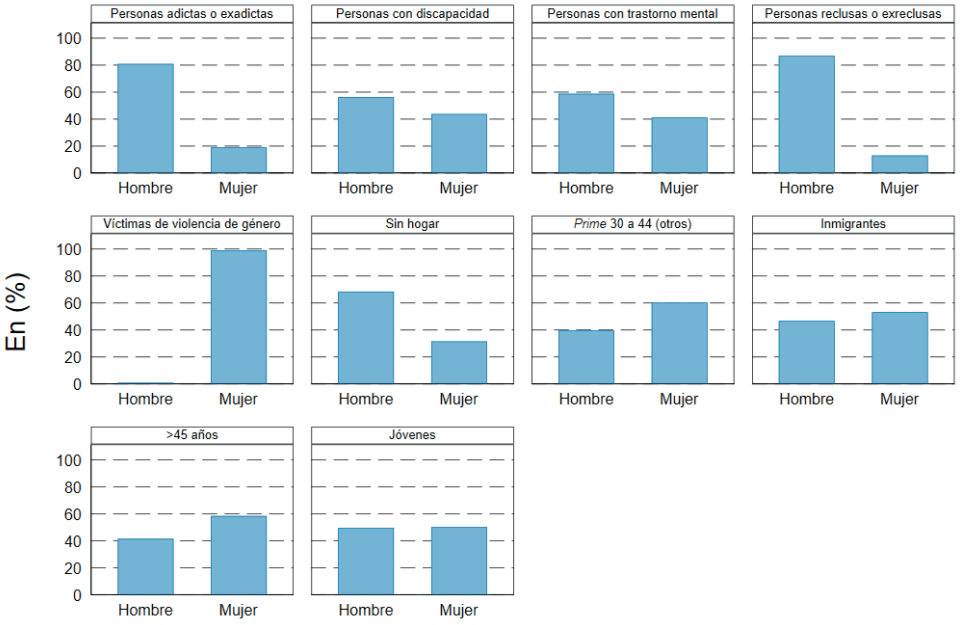
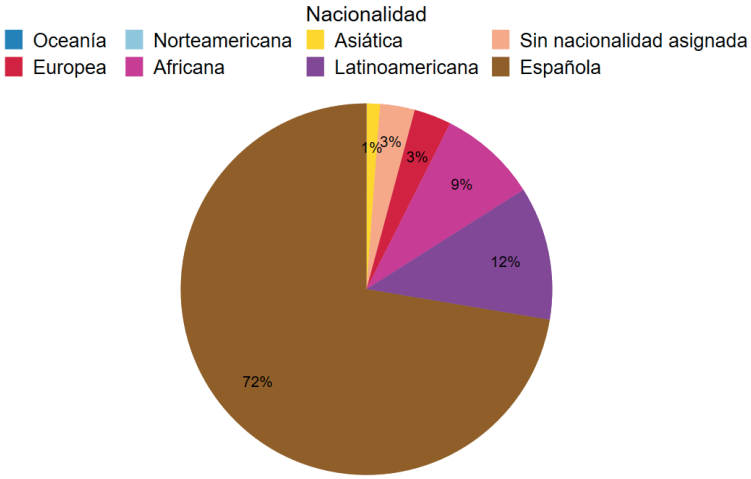
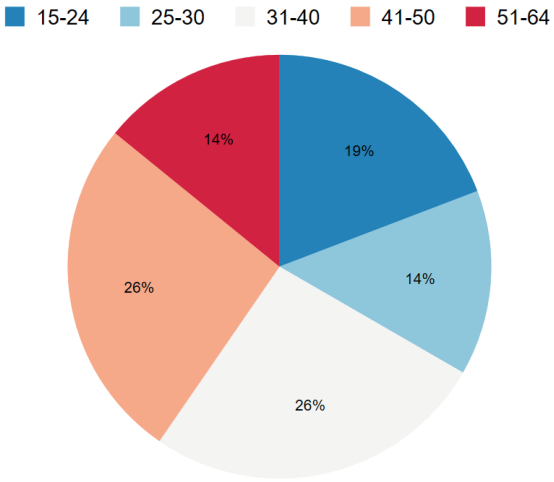


FIGURA 2 *(Continuación)*
DESCRIPTIVOS SOBRE LOS BENEFICIARIOS

c) Porcentaje de beneficiarios por nacionalidad

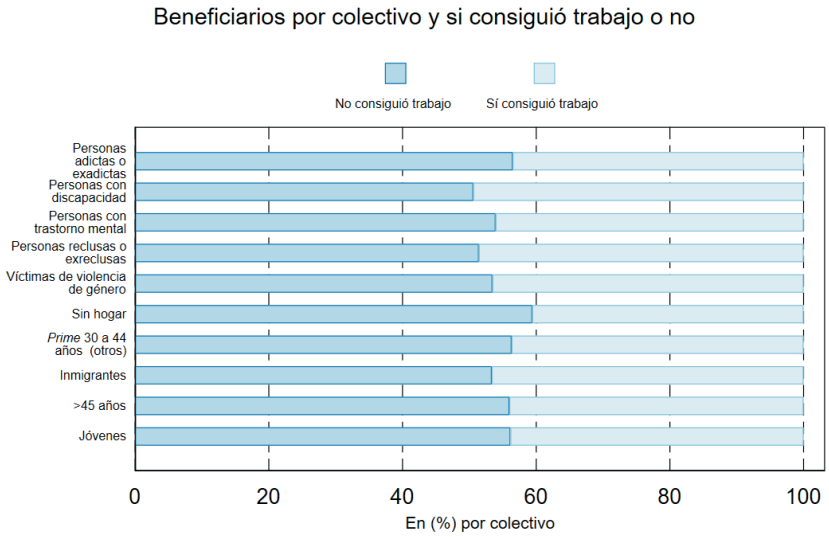


d) Porcentaje de beneficiarios por edad



FUENTE: Elaboración propia en base a datos programa Incorpora.

FIGURA 3
DESCRIPTIVOS SOBRE LOS BENEFICIARIOS Y EL PORCENTAJE DE EMPLEADOS



NOTA: La información sobre el estado laboral de todos los participantes proviene de la plataforma Incorpora bajo el estado «Trabajo por cuenta ajena» o «Trabajo por cuenta propia». Condicionamos a tener trabajo si se encuentra más de un mes empleado tras pasar por el programa.
FUENTE: Elaboración propia en base a datos programa Incorpora.

3.2. Metodología

Una vez construida la base de datos estimamos un modelo econométrico de empleabilidad para los beneficiarios del programa, teniendo en cuenta la posible endogeneidad en la asignación de formación al participante en el programa.

Una de las características de Incorpora es la posibilidad de ofrecer cursos de formación para mejorar los conocimientos y habilidades de los beneficiarios. Esta asignación no es aleatoria y depende, no solo de sus características, sino también del técnico que los evalúa y la entidad en la que se registra. Por tanto, una dificultad del análisis es determinar aquella información que permita corregir la posible endogeneidad en la asignación de la formación.

Otra de las dificultades para evaluar la formación es que, durante ese tiempo, el beneficiario puede o no estar trabajando y el efecto que observemos no sea homogéneo, sino que deba desagregarse.

Resulta necesario, en ambos casos, incluir la información específica sobre el técnico que acompaña al beneficiario y el tipo de formación que realiza para poder cuantificar el efecto de esta.

El modelo con el que se desarrolla el perfilado es un modelo econométrico de elección discreta, concretamente un modelo Probit. Siguiendo la metodología propuesta por Felgueroso *et al*. (2018), la variable dependiente es una variable binaria que adopta un valor de 1, si el beneficiario transita del desempleo al empleo antes del periodo *T*; y de 0, en caso contrario.

La expresión analítica del modelo a estimar en este caso es:

$$\Pr[\textit{Salida al empleo} \mid x] = \Phi(X'\beta + \varepsilon) \quad \text{con } \varepsilon \sim N(0,1)$$

donde Φ denota una función de probabilidad y el término de error tiene una distribución normal estándar de varianza normalizada a 1.

Las variables de control (X) que se consideran son las siguientes:

– *Características de los beneficiarios:* edad, nacionalidad, colectivo, número de hijos, nivel de formación, sexo, grado de discapacidad, ocupación solicitada, disponibilidad para viajar vehículo propio, idioma, situación laboral, proximidad del empleo deseado, tipo de jornada deseada y efecto fijo de técnico de atención en el programa.
– *Variable de formación:* cursos cortos (menos de 3 meses de duración) y cursos largos (más de 3 meses). Cada una representa una variable discreta con tres etapas: sin formación, durante la formación y después de la formación.
– *Características de las entidades:* tipo de entidad (de red o coordinadora), servicios que ofrece (punto formativo, punto reincorpora, punto retorno educativo, Servicio comunidad, punto auto-ocupación, Fondo Social Europeo y/o punto Incorpora joven), tipo de colectivo principal y secundario.
– *Características del mercado Incorpora:* número de ofertas de empleo (por entidad compartida y mes), desagregadas entre duración del contrato (1 semana, 1 mes, 3 a 12 meses y 1 a 3 años), número de horas (al menos 40 horas o menos), colectivo de preferencia (discapacitado, en riesgo de exclusión social o indiferente), sector de actividad (industria, comercio, agricultura o servicios).
– *Características de las empresas inscritas:* Número de empresas inscritas en Incorpora por mes, sector y si tienen o no convenio firmado.
– *Características temporales:* año y mes.
– *Características de las acciones:* duración del paro, duración hasta la formación, si tiene o no formación en el episodio previo y si tiene o no un empleo previo.

Siguiendo, en ese caso, la siguiente especificación:

$$X = X'_{\text{benef}}\beta_{\text{benef}} + X'_{\text{formación}}\beta_{\text{formación}} + X'_{\text{entidad}}\beta_{\text{entidad}} + X'_{\text{mercado}}\beta_{\text{mercado}} + X'_{\text{empresas}}\beta_{\text{empresas}} + Z'\alpha + X'_{\text{acciones}}\beta_{\text{acciones}}$$

Y, en el caso de formación, tenemos que:

$$X'_{\text{formación}}\beta_{\text{formación}} = \sum_{J=1}^{2} X_{\text{durante } j}\beta_{\text{durante } j} + X_{\text{después } j}\beta_{\text{después } j} +$$

$$+ \sum_{i=1}^{9} \text{mes } i \; x \; X_{\text{después } j \; x \text{ mes } i}\beta_{\text{después } j \; x \text{ mes } i}$$

donde $j = 1$ si el curso es corto y $j = 2$ si es largo.

Una vez estimado el modelo, con una determinada periodicidad, la predicción para cada uno de los individuos se define como el grado de empleabilidad de dicho individuo (*empleabilidad* = $\Phi(X'\hat{\beta})$)

La *probabilidad estimada de salida del paro al empleo en un mes* para cada beneficiario se puede considerar como un indicador de empleabilidad individual que se utiliza a continuación con los siguientes propósitos:

En primer lugar, clasificar a los beneficiarios en grupos de empleabilidad y así poder analizar las características de los beneficiarios incluidos en cada grupo. La clasificación propuesta aquí se realiza en función de la pertenencia a uno de los cuartiles de las distribuciones de la probabilidad de salir del paro al empleo. Así, establecemos cuatro grupos: empleabilidad alta, media alta, media baja y baja. Esto nos permite ver cómo están repartidas las características de los beneficiarios entre cada umbral e identificar aquellas personas que presentan mayores dificultades para encontrar un empleo.

En segundo lugar, estimar, entre otros posibles factores, el efecto neto de la formación y cómo este varía en el tiempo. Recurrimos al efecto marginal promedio para cuantificar el efecto durante y después de la formación a partir de la siguiente fórmula:

$$\textit{Efecto marginal promedio} = \frac{1}{n} \sum_{i=1}^{n} [\Phi(x'_i\beta \mid x_i^k = 1) - \Phi(x'_i\beta \mid x_i^k = 0)]$$

Consideramos tres posibles tipos de estado respecto a la formación:

- Antes: si el beneficiario no ha realizado ningún curso.
- Durante: si en el periodo *t* se encuentra realizando la formación.
- Después: una vez finalizado el curso y hasta su censura o baja del programa.

En una fase ulterior, podemos analizar dicho efecto para cada colectivo, y dentro de estos, por género o grupo de edad, para tomar en cuenta la heterogeneidad de beneficiarios que hay.

3.3. Resultados

A continuación, se presentan los principales resultados del modelo previamente descrito. Todas las regresiones incluyen los controles previamente señalados (efectos fijos sobre características demográficas, variación temporal (año y mes), características y herramientas del programa Incorpora (cursos, número de ofertas, tipos de entidad, entre otros) y características previas a la entrada en Incorpora (vida laboral previa, situación laboral). En primer lugar, en el Cuadro 1 presentamos los efectos marginales referidos a las variables de formación ($X'_{\text{formación}}$).

CUADRO 1

EFECTOS MARGINALES DE LOS CURSOS DE FORMACIÓN SOBRE LA REGRESIÓN DE PERFILADO

Variables	Efectos marginales	Variables	Efectos marginales
Cursos cortos		**Cursos largos**	
Durante	0,0637*** (0,00058)	Durante	0,0250*** (0,00116)
Después	0,00517*** (0,000532)	Después	0,00620*** (0,00142)
Después × mes 1	0,0221*** (0,00111)	Después × mes 1	0,0171*** (0,00338)
Después × mes 2	0,00562*** (0,00114)	Después × mes 2	–0,0159*** (0,00347)
Después × mes 3	–0,0179*** (0,00115)	Después × mes 3	–0,0123*** (0,00352)
Después × mes 4	–0,0197*** (0,00116)	Después × mes 4	–0,0169*** (0,0036)
Después × mes 5	–0,0155*** (0,00118)	Después × mes 5	–0,0177*** (0,00367)
Después × mes 6	–0,0122*** (0,00121)	Después × mes 6	–0,00882** (0,00376)
Después × mes 7	–0,0146*** (0,00123)	Después × mes 7	–0,0117*** (0,00387)
Después × mes 8	–0,0184*** (0,00126)	Después × mes 8	–0,0182*** (0,00395)
Después × mes 9	–0,0150*** (0,00128)	Después × mes 9	–0,0141*** (0,00405)
Observaciones	4.285.429	Observaciones	4.285.429
R-cuadrado	0,05	R-cuadrado	0,05

NOTA: Errores estándares en paréntesis. * p < 0,10. ** p < 0,05. *** p < 0,01. Los regresores omitidos en la tabla incluyen características demográficas (X'_{benef}), de las entidades (X'_{entidad}), de las ofertas de empleo publicadas en Incorpora (mercado Incorpora), de las empresas inscritas en el programa (X'_{empresas}), *dummies* de año y mes (Z) y características de las acciones de los beneficiarios (X'_{acciones}). El listado más detallado se encuentra en la sección 3.2.

FUENTE: Elaboración propia.

El efecto de la formación sobre la probabilidad de salir del paro resulta significativo tanto durante como después de haberla percibido, para todos los beneficiarios del programa considerados en el periodo 2011-2019. Este resultado es importante ya que indica que la formación no se considera un periodo de inactividad, sino que mejora las oportunidades de empleo. Esto tiene que ver con la orientación de los cursos, ya que en muchos casos ofrecen herramientas relacionadas con la búsqueda de trabajo.

En la Figura 4 presentamos un resumen de los resultados respecto a la formación: efecto marginal durante el periodo formativo y después del mismo. Asimismo, representamos la evolución del efecto marginal en el tiempo, tanto para cursos cortos como para cursos largos. En este último caso, también presentamos los resultados según el colectivo estudiado.

A partir de los datos suministrados por la plataforma Incorpora, se estimó que los cursos de formación aumentan las probabilidades de encontrar un empleo entre un 3 y un 6 % adicional, siendo los de menor duración aquellos que mayor efecto tienen. Este, además, se observa desde el momento en que el beneficiario inicia la formación, ya que el hecho de asistir parece tener un impacto inmediato sobre su empleabilidad. No obstante, una vez finalizada la formación las ganancias del curso disminuyen conforme transcurren los meses.

FIGURA 4
EFECTO DE LA FORMACIÓN SOBRE LA SALIDA DEL PARO

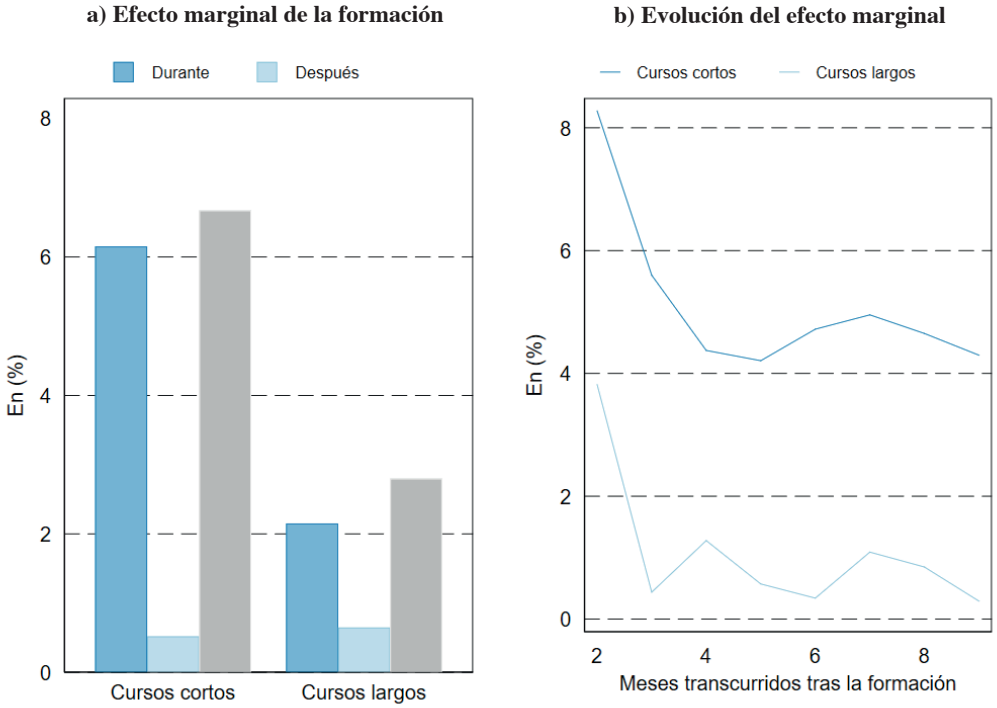


FIGURA 4 *(Continuación)*
EFECTO DE LA FORMACIÓN SOBRE LA SALIDA DEL PARO

c) Efectos de la formación

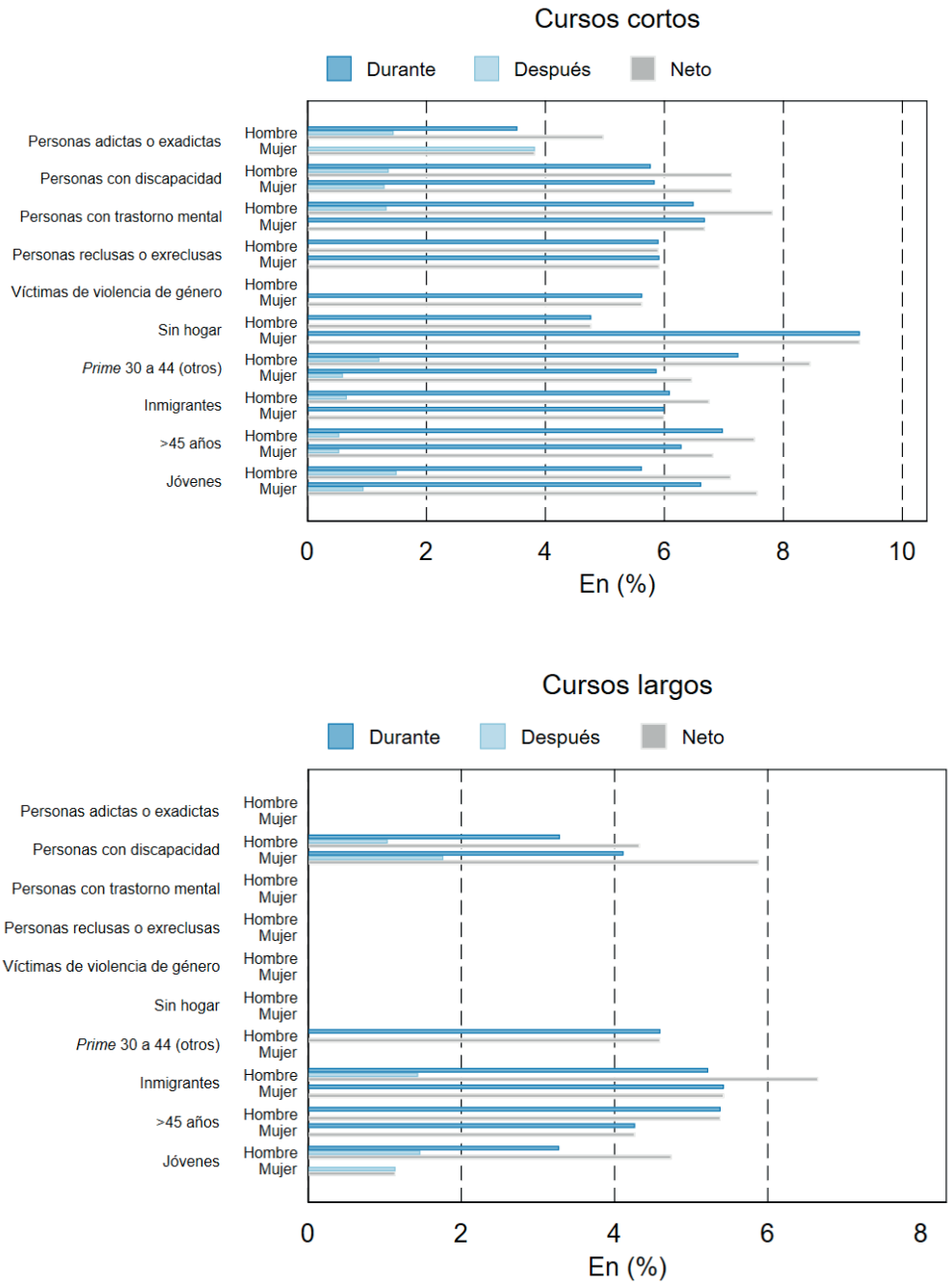
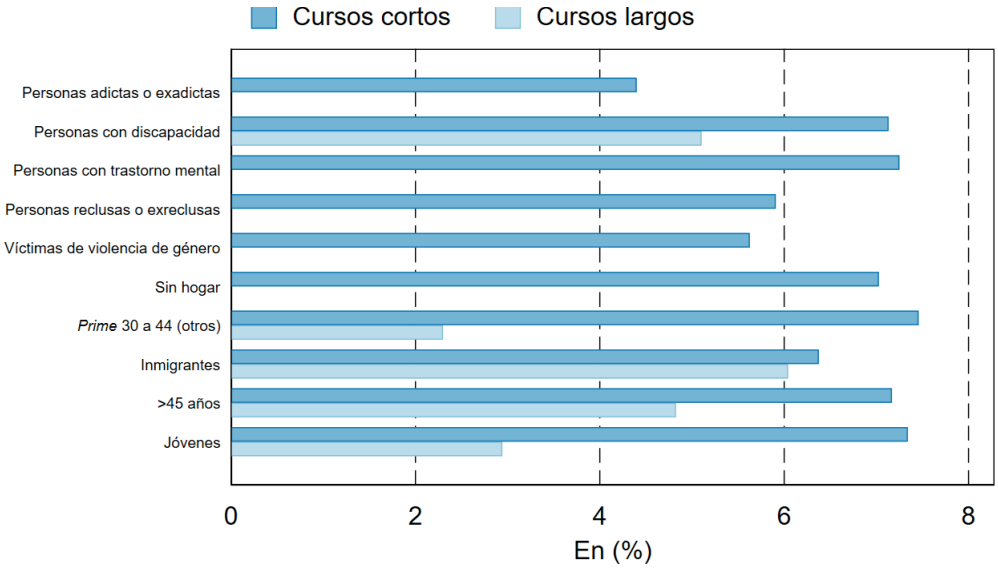


FIGURA 4 ***(Continuación)***
EFECTO DE LA FORMACIÓN SOBRE LA SALIDA DEL PARO

d) Efecto marginal de la formación sobre la empleabilidad (según el colectivo)



FUENTE: Elaboración propia en base a datos programa Incorpora.

Por otra parte, es importante analizar si estas diferencias también se ven reflejadas en la distribución de la empleabilidad. En la Figura 5 presentamos la distribución de empleabilidad mensual (probabilidad de dejar el no empleo en un mes) para el conjunto de beneficiarios. En el conjunto de los beneficiarios la gran masa de probabilidad se encuentra entre el 0 y el 13 %, estando la moda de la distribución muy a la derecha, entre el 1 y el 2 %. El percentil 50 en algo menos del 6 % y el percentil 75 entorno al 8 %.

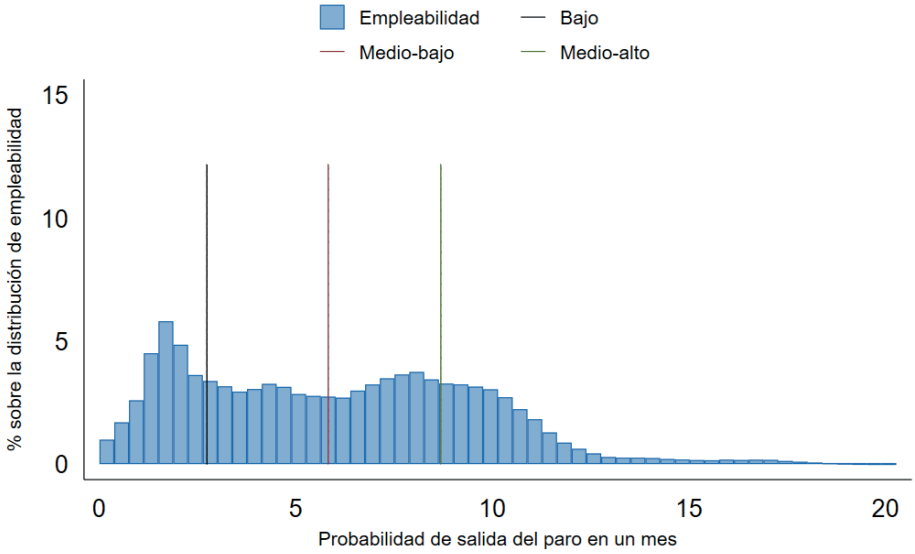
En segundo lugar, comparamos la empleabilidad de los beneficiarios al momento de inscribirse en el programa. El Cuadro 2 señala que las probabilidades de salida del paro en un mes más altas son las de los jóvenes (6,61 puntos) y las menores las de las personas adictas o exadictas (4,53 puntos). En términos de dispersión, detectamos que la diferencia entre el primer y el tercer cuartil oscila entre 5,7 (jóvenes) y 6,15 (inmigrantes) puntos de probabilidad. Solo para las personas sin hogar la dispersión es sensiblemente menor (solo 4,6 puntos).

La Figura 6 señala importantes diferencias en la distribución de aquellos con y sin formación. Mientras que la media se encuentra entorno al 6 % para aquellos que no la percibieron, esta es superior (8 %) para el caso contrario.

Finalmente, una de las preguntas más relevantes para los coordinadores del programa es si los recursos en términos materiales y de capital humano se están

FIGURA 5

DISTRIBUCIÓN DE LA EMPLEABILIDAD PROMEDIO DE TODOS LOS BENEFICIARIOS



NOTA: Las rectas indican los cuartiles estimados.
FUENTE: Construcción propia en base a datos programa Incorpora.

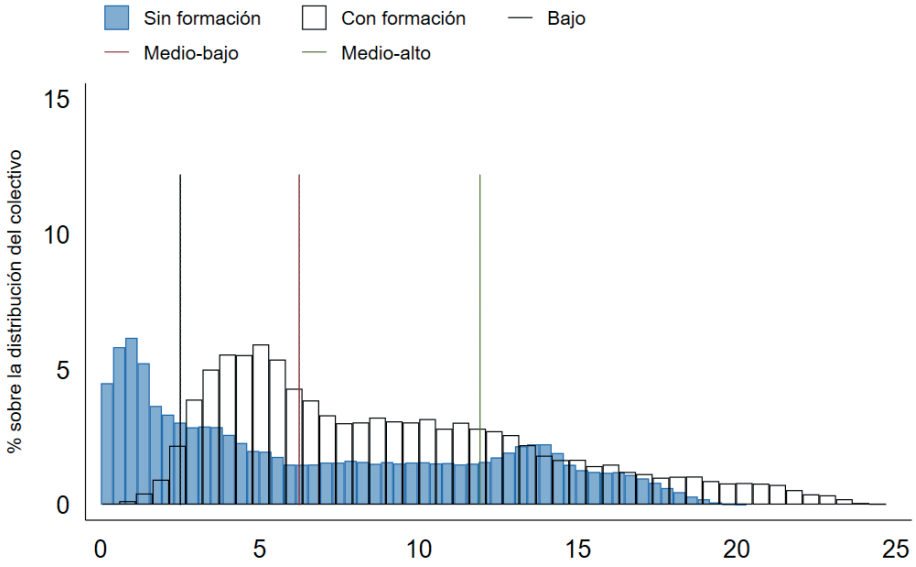
CUADRO 2

PERCENTILES DE EMPLEABILIDADES INICIALES POR COLECTIVO

Probabilidad de salida del paro en un mes	**Percentil**		
Colectivo	**25**	**50**	**75**
Personas adictas o exadictas	2,03	4,53	7,83
Personas con discapacidad	2,07	4,56	7,89
Personas con trastorno mental	2,14	4,87	7,98
Personas reclusas o exreclusas	2,80	6,14	8,81
Víctimas de violencia de género	2,65	5,56	8,31
Sin hogar	2,19	4,56	6,86
Prime (otros) 30 a 44 años	2,94	6,01	8,77
Inmigrantes	2,87	6,03	9,02
Mayores de 45 años	3,14	6,24	8,92
Jóvenes	3,47	6,61	9,11

FUENTE: Elaboración propia.

FIGURA 6
DISTRIBUCIÓN DE LA EMPLEABILIDAD SEGÚN LA FORMACIÓN RECIBIDA

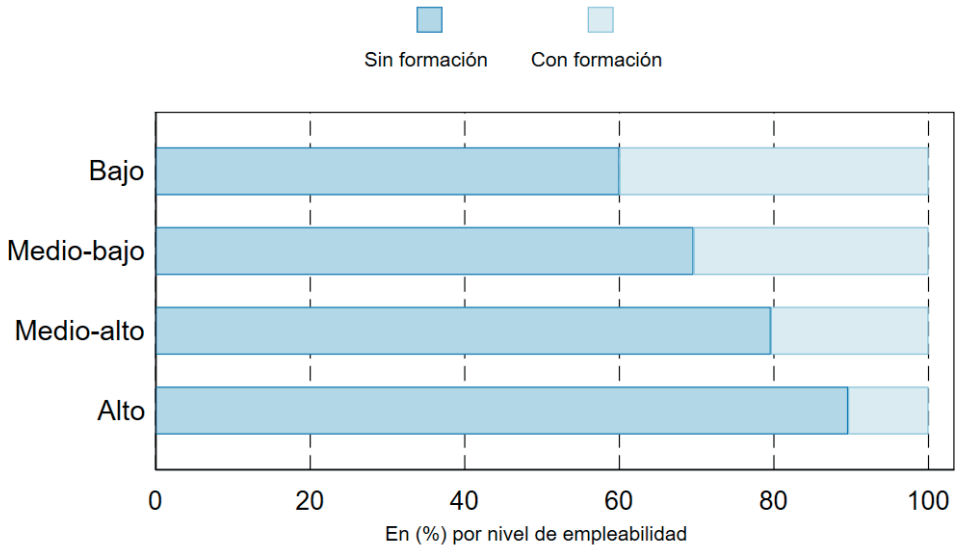


NOTA: Las rectas marcan los límites de la empleabilidad en general.
FUENTE: Elaboración propia en base a datos programa Incorpora.

aportando a aquellos que más lo necesitan. La Figura 7 muestra la distribución de la formación según el nivel de empleabilidad que tienen los beneficiarios antes de recibir dicha formación. Si bien los beneficiarios con una baja empleabilidad que reciben formación representan un porcentaje relativamente alto, también ocurre con aquellos cuyas probabilidades de salir del paro al momento de acceder a Incorpora son relativamente buenas. Esto podría mejorarse, si se enfocara el programa más en aquellos beneficiarios que más necesitan la formación como medio para aumentar su empleabilidad. Y más aun habiendo demostrado que efectivamente los cursos generan ganancias significativas sobre las probabilidades de encontrar un empleo.

En general encontramos, por tanto, efectos positivos de la formación, más importantes para los cursos cortos que para los largos. Los efectos de los primeros son importantes para prácticamente todos los colectivos estudiados. En cambio, los segundos solo encontramos que son significativos para los colectivos más importantes (discapacitados, inmigrantes, 30-45, 45+ y jóvenes).

FIGURA 7
EL EFECTO DE LA FORMACIÓN SEGÚN EL CUARTIL DE EMPLEABILIDAD
Beneficiarios por nivel de empleabilidad inicial y formación percibida en Incorpora



FUENTE: Elaboración propia en base a datos programa Incorpora.

4. Análisis causal

Hasta ahora, el análisis solo permitía comparar los beneficios del programa entre colectivos, pero no estudiar su impacto. Para ello, incorporamos en esta sección una base de datos con la vida laboral de los beneficiarios del programa. Esto nos permite analizar causalmente el efecto de Incorpora a corto y medio plazo, construyendo un grupo de control y cuantificando el efecto promedio entre los tratados.

4.1. Datos

La información para realizar el análisis causal está basada en la Muestra Continua de Vidas Laborales (MCVL)[2], que es un extracto de datos individuales anonimizados, procedentes de las bases de datos de la Seguridad Social y a los que se añaden otros que se toman del Padrón Continuo Municipal (INE) (García Pérez, 2008). En nuestro caso, esta muestra se refiere a las vidas laborales de una muestra aleatoria de beneficiarios del programa Incorpora y que no forman parte de la MCVL de libre acceso. En el

[2] www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST211

Cuadro 3 se puede observar la distribución de los beneficiarios según el tipo de colectivo, siendo los jóvenes aquel con un mayor peso y las personas adictas con el menor[3].

Una vez depurados los datos, nos quedamos con la siguiente información de tratados por el programa Incorpora.

CUADRO 3
FRECUENCIAS SOBRE EL NÚMERO DE BENEFICIARIOS CON DATOS DE LA SEGURIDAD SOCIAL

Colectivo	Freq.	Percent
Personas adictas o exadictas	28	0,03
Personas con discapacidad	15.535	19,29
Personas con trastorno mental	2.951	3,66
Personas reclusas o exreclusas	1.997	2,48
Víctimas de violencia de género	1.111	1,38
Sin hogar	146	0,18
Prime 30 a 44 (otros)	15.883	19,72
Inmigrantes	12.490	15,51
>45 años	12.187	15,13
Jóvenes	18.203	22,6

FUENTE: Elaboración propia.

Esta información luego es cruzada con un *pool* de Muestras Continuas de Vidas Laborales (2007-2018) donde tenemos información equivalente a la que tenemos sobre tratados de los potenciales controles[4].

En este punto es conveniente destacar que restringimos el análisis causal a los colectivos que podemos identificar suficientemente bien en la muestra de control. Concretamente, el análisis se realizará para los siguientes colectivos: personas con discapacidad, *prime* (30-44), jóvenes, inmigrantes y mayores de 45 años. Otros colectivos requerirían el cruce con bases de datos adicionales (de salud, penitenciarias, o sociales) para poder recolectar información de población equivalente que no haya participado en el programa. Por esta, y no otra razón, los mantenemos al margen del análisis causal presentado en el presente artículo.

[3] Condicionar el análisis a la presencia de los trabajadores en los datos de la Seguridad Social podría generar un sesgo de selección, ya que obvia a aquellos que acceden al mundo laboral gracias al programa y no encuentran empleo (Heckman, 1979). No obstante, no excluimos a estos individuos de la muestra y tomamos en cuenta su vida laboral.

[4] Es posible que existan beneficiarios del programa Incorpora cuya vida laboral se encuentre en la MCVL. Para excluir estos casos, emparejamos individuos en ambas bases de datos usando sus datos personales y de afiliación. El resultado es que 23 individuos se encuentran en ambas bases datos y fueron excluidos de la base de datos de la MCVL para evitar duplicados.

4.2. Metodología

Para analizar el efecto causal del programa Incorpora, es necesario crear antes una muestra de control comparable a la muestra de tratados y que permita evaluar el efecto del tratamiento en los tratados (*average treatment effect over the treated*, ATT).

El problema surge en el momento de crear dicha muestra de control. Si todos los beneficiarios se dieran de alta al mismo tiempo, existiría un periodo de tratamiento consistente a partir del cual es posible crear una muestra única. No obstante, las altas en Incorpora se producen de manera sucesiva desde enero del 2011 hasta el día de hoy y, por tanto, resulta necesario crear muestras que tengan características similares no solo inherentes (demográficas), sino también en el momento de evaluarse.

Por ello, nuestro análisis procede en primer lugar a separar por colectivos objeto de estudio y años de alta en el programa a los tratados y a la muestra que en cada caso se usa como control[5]. En el caso de los potenciales controles, se consideran, además, aquellos que cumplan con unas características demográficas previas:

– Discapacitados: personas que tengan un grado de discapacidad.
– Jóvenes: personas que tengan entre 16 y 29 años entre 2011 y 2016, y no sean discapacitadas ni inmigrantes.
– Mayores de 45 años: personas que tengan más de 45 años entre 2011 y 2016, y no sean discapacitadas ni inmigrantes.
– *Prime* (Otros): personas que tengan entre 30 y 44 años entre 2011 y 2016, y no sean discapacitadas ni inmigrantes.

Para cada año, consideramos como punto de referencia el mes previo al alta en Incorpora (tratados) o diciembre del año previo para los controles (como ejemplo, intentaremos crear un grupo de control para los beneficiarios del 2011 con la información de los potenciales controles en diciembre del 2010).

4.2.1. Primera etapa: *propensity score matching*

El primer caso consiste en crear dicho grupo de control a partir de técnicas del *propensity score matching* (Rosenbaum & Rubin, 1983), donde se estima primero el siguiente modelo Probit:

$$\Pr(T1 = 1|Z) = \Phi(Z'\beta)$$

Siendo $T1$ la variable de tratamiento (o participación del programa Incorpora en este caso) y Z las variables de control. Los regresores Z van a condicionar el

[5] En este trabajo analizamos el efecto promedio del programa sobre aquellos que cumplen dichas características y pueden ser emparejados con la muestra. Un análisis más detallado y separado por colectivos aparecerá publicado próximamente por la Fundación Obra Social La Caixa.

análisis ya que, según la información que usemos, el modelo resultará en un grupo de tratamiento y control distinto. En nuestro criterio base[6], construimos las regresiones siguiendo lo que a partir de ahora denominaremos «criterio Incorpora», esto es, basaremos la selección en información que los técnicos registran en la base de datos de Incorpora pero que también está en los ficheros de la Seguridad Social. En concreto, las variables que incluimos son:

$$Z_i = edad_alta_i + género_i + educa_i + 1\{prestación_i\} + meses_paro_i$$

Donde $edad_alta_i$ corresponde a la edad de la primera alta laboral, $género_i$ a una dicotómica de hombre o mujer, $educa_i$ a siete niveles de nivel educativo (sin estudios, primaria, secundaria, FP, superior, estudios no reglados y no consta), indicador de si percibió o no prestación por desempleo en el último empleo, y número de meses de desempleo ($meses_paro_i$).

Al mismo tiempo, se añade información adicional dependiendo del tipo de colectivo. Así, para los inmigrantes, la nacionalidad está en el emparejamiento, mientras que para los discapacitados se incluye el grado de la discapacidad.

Para cada año (2011-2015) se estima un modelo con 3 años pre y 3 años postratamiento, considerando únicamente a beneficiarios del programa dados de alta en ese año (ejemplo: para los dados de alta en 2011, consideramos el periodo 2008-2014).

El procedimiento del *propensity score matching* es el siguiente:

En primer lugar, se estima el modelo Probit con las variables de control mencionadas y se hace una predicción de la probabilidad de ser un tratado. Luego, con estos datos se estima una densidad de kernel para representar la distribución de la predicción y se empareja a cada tratado con aquel control cuya probabilidad sea lo más parecida posible usando la siguiente expresión:

$$K_i = K\left(\frac{p(x_i) - p(x_k)}{h_n}\right)$$

Donde h es la banda (*bandwidth)* seleccionada y K es el kernel de Epanechnikov (1969). De esta forma, se obtiene un control con las mismas probabilidades de ser elegido. Un elemento importante de mencionar es que los controles no son excluyentes entre años, es decir, un control para los tratados del año t también puede serlo luego para los tratados del año $t + 1$.

Obtenido el grupo de control para cada colectivo y año de análisis, procedemos a analizar el efecto causal de Incorpora sobre resultados en el mercado de trabajo de los diferentes colectivos considerados en Incorpora.

[6] Están disponibles, a petición del lector interesado, resultados con criterios alternativos, más estrictos, que validan los resultados obtenidos en el escenario base.

4.2.2. Segunda etapa: análisis de impacto

En esta sección evaluamos causalmente el impacto del tratamiento Incorpora en diversas variables de impacto en base al modelo de regresión. Todas las regresiones se estiman por Mínimos Cuadrados Ordinarios[7] en base a:

$$Y_{it} = \beta_0 + \beta_1 T1\ Incorpora_{it} * Post_{it} + \beta T\, X_{it} + \alpha_i + \alpha_t + \varepsilon_i, \quad i = 1, 2, ..., n$$

Donde Y es la variable objeto de análisis, *T1 Incorpora* es una variable ficticia que vale 1 para los beneficiarios emparejados en la primera etapa, X son todas las variables de control, α_i son efectos fijos de individuo, α_t son efectos fijos de tiempo, ε el residuo y la unidad de tiempo t es trimestral[8].

Se plantearon tres preguntas principales que corresponden a variables dependientes (Y) diferentes, o a valores del parámetro β distintos, esto es, a efectos del tratamiento definidos de manera diferente:

1. ¿Cuál es el efecto del programa Incorpora sobre la probabilidad de salir del paro?

$$Y_{1i} = \begin{cases} 1, & si\ en\ t+1\ el\ individuo\ i\ sale\ del\ paro \\ 0, & si\ en\ t\ se\ encuentra\ parado \end{cases}$$

2. ¿Cuál es el efecto del programa Incorpora sobre el salario posterior a Incorpora?

$$Y_{2i} = \ln(base\ de\ cotización)_{it}\ del\ individuo\ i\ en\ el\ trimeste\ t$$

3. ¿Cuál es el efecto del programa Incorpora sobre el número de meses acumulados en el empleo?

$$Y_{3i} = Meses\ acumulados\ en\ el\ empleo\ desde\ el\ alta\ en\ Incorpora\ a\ n\ trimestres$$

Para todos los casos, los regresores (X_{it}) se mantienen constantes y son los siguientes:

– Edad de la primera alta.
– Género.

[7] Los modelos de probabilidad lineal (MPL) se definen como aquellos modelos de decisión discreta donde la estimación se hace a través de Mínimos Cuadrados Ordinarios. Estos modelos tienen ventajas y desventajas: por una parte, la estimación es menos exigente (no es el caso del Probit, donde se estiman los coeficientes por Máxima Verosimilitud) y los coeficientes son muy similares a los estimados en un modelo con distribución normal o logística; por otra, las magnitudes del valor promedio de Y_{it} para casos extremos son difíciles de interpretar ya que son Y_{it} mayores a 1 o menores a 0.

[8] Este modelo sigue una especificación Twoway Fixed Effects DD con variación temporal en el tratamiento (Goodman-Bacon, 2018).

– Nivel educativo en 7 niveles.
– CC. AA. de la última cuenta de cotización.
– Indicador de si percibió o no prestación por desempleo en el último desempleo.
– Indicador de si percibió o no prestación por discapacidad en el último desempleo.
– Sector de actividad económica (CNAE) del último empleo.
– Número de meses parados desde el último empleo.
– Último tipo de contrato.

Mientras que en el modelo que estima el impacto en el salario se añaden, además, los siguientes regresores referidos a la empresa donde se encuentra el empleo:

– Sector del empleo de alta.
– Número de trabajadores en la cuenta de cotización (6 tramos que van desde 1 hasta > = 1.500 trabajadores).

4.3. Resultados

En primer lugar, presentamos el número de observaciones resultantes de la primera etapa. El método de emparejamiento de Kernel nos permite asignar más de un control a un tratado como se puede observar en el primer Cuadro 4. Además, el número de controles es, en líneas generales, constante para cada periodo de análisis.

Más interesante aún es el test de medias que elaboramos para analizar posibles diferencias en términos de características observables (Rosenbaum & Rubin, 1985). No existen diferencias importantes entre ambos grupos y para ninguno de los años, lo que parece indicar que el emparejamiento es bueno y permite inferir causalidad ante la presencia de una formación como Incorpora. Cabe recordar que en ningún momento estimamos el efecto promedio del programa (ATE) sino el efecto de Incorpora entre los tratados (ATT), ya que la decisión de participar es voluntaria.

En segundo lugar, una vez realizada la segunda etapa encontramos que el efecto del programa sobre la probabilidad de salir del desempleo es positivo y creciente con el año estudiado (Cuadro 5). Empieza en algo más de cuatro puntos de probabilidad y termina en nueve. Asimismo, son significativos, importantes y crecientes en el tiempo los efectos encontrados sobre la probabilidad de estar empleado a uno, cuatro, ocho y once trimestres vista.

En el promedio también encontramos efectos significativos sobre la base reguladora asociada al empleo encontrado y el número de meses en el empleo acumulados en todos los horizontes analizados. Esto se debe a unos errores estándares relativamente menores respecto al resto de colectivos, lo que supone que la inferencia tenga mayores probabilidades de indicar efectos significativos.

En definitiva, en el corto y medio plazo el programa Incorpora ocasiona un impulso efectivo en la empleabilidad del conjunto de los colectivos estudiados.

CUADRO 4

NÚMERO DE OBSERVACIONES Y CUADRO DE MEDIAS PARA EL PROMEDIO DE TODOS LOS BENEFICIARIOS ANALIZADOS

	***Matching* de criterios Incorpora**					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Controles	439.988	439.555	439.700	439.479	439.849	440.355
Tratados	3.173	4.924	7.875	9.117	10.903	14.425
	Medias					
	2011		**2012**		**2013**	
	Tratados	**Controles**	**Tratados**	**Tratados**	**Controles**	**Controles**
Edad de la primera alta	23,478	30,316	23,355	30,321	22,758	30,315
Últimos meses paro	10,996	11,309	13,864	15,061	16,777	18,827
Mujeres	0,469	0,485	0,499	0,485	0,510	0,485
Sin estudios	0,031	0,133	0,023	0,133	0,022	0,133
Primaria	0,256	0,157	0,239	0,157	0,233	0,156
Secundaria	0,292	0,235	0,289	0,235	0,289	0,235
FP	0,178	0,168	0,222	0,168	0,228	0,168
Superior	0,102	0,125	0,130	0,125	0,146	0,125
Estudios no reglados	0,141	0,180	0,097	0,181	0,083	0,182
	2014		**2015**		**2016**	
	Tratados	**Controles**	**Tratados**	**Controles**	**Tratados**	**Controles**
Edad de la primera alta	22,766	30,318	22,922	30,304	23,031	30,278
Últimos meses paro	21,425	22,648	27,627	26,464	32,750	30,084
Mujeres	0,502	0,485	0,497	0,485	0,505	0,486
Sin estudios	0,019	0,134	0,024	0,133	0,025	0,133
Primaria	0,232	0,157	0,249	0,157	0,262	0,158
Secundaria	0,308	0,235	0,327	0,236	0,305	0,235
FP	0,231	0,167	0,219	0,168	0,225	0,168
Superior	0,136	0,126	0,121	0,125	0,138	0,125
Estudios no reglados	0,073	0,182	0,060	0,181	0,045	0,181

FUENTE: Elaboración propia.

CUADRO 5
RESULTADOS PRINCIPALES PARA EL PROMEDIO DE TODOS LOS BENEFICIARIOS ANALIZADOS

Efectos sobre la salida del paro						
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	0,04148*** (0,00372)	0,04924*** (0,00301)	0,06021*** (0,00258)	0,06780*** (0,00253)	0,07426*** (0,00239)	0,09631*** (0,00199)
Observaciones	5.076.284	5.233.094	5.424.347	5.525.972	5.592.003	5.604.019
	Efectos sobre los salarios					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	0,03541 (0,02443)	0,02795 (0,01838)	0,03786** (0,01544)	0,05955*** (0,01513)	0,04881*** (0,01592)	0,07134*** (0,01345)
Observaciones	306.579	292.748	303.907	315.361	328.479	343.453
Efectos sobre meses acumulados en el empleo						
	Meses acumulados a 1 trimestre					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	1,78033*** (0,17712)	1,55174*** (0,14883)	1,67671*** (0,10127)	1,81286*** (0,11342)	1,78590*** (0,10096)	1,97372*** (0,09249)
Observaciones	442.582	443.634	446.425	447.446	449.490	453.562
	Meses acumulados a 4 trimestres					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	3,26881*** (0,59846)	2,25714*** (0,44261)	3,11020*** (0,28008)	3,08167*** (0,33266)	3,13656*** (0,28039)	3,34951*** (0,25066
Observaciones	442.431	443.317	445.892	446.864	448.708	452.557
	Meses acumulados a 8 trimestres					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	3,87795*** (1,16215)	2,70467*** (0,8273)	4,39270*** (0,5066)	3,74612*** (0,60294)	4,79458*** (0,53755)	4,94722*** (0,46414)
Observaciones	441.546	441.901	443.482	443.981	445.366	448.085
	Meses acumulados a 11 trimestres					
	2011	**2012**	**2013**	**2014**	**2015**	**2016**
Tratados Incorpora	2,75309* (1,52591)	1,60359 (1,072)	4,27771*** (0,70467)	3,39323*** (0,83186)	4,40315*** (0,76505)	4,86818*** (0,69351)
Observaciones	440.039	439.642	439.921	439.642	440.060	440.612

FUENTE: Elaboración propia.

Los resultados indican que, en el promedio de beneficiarios analizados, el programa Incorpora tiene un efecto positivo sobre la probabilidad de salida del paro una vez dentro del programa, positivo sobre el número de meses acumulados en el empleo y sobre los salarios (aunque este último desaparece cuando se analiza colectivo a colectivo). Asimismo, cuando analizamos los resultados para cada colectivo por separado encontramos efectos algo más importantes para el grupo de inmigrantes (entre un 4,8 % y un 12 %) y para los discapacitados (entre un 5,9 % y un 6,9 %), que para los otros tres grupos poblacionales donde distinguimos por tramos de edad. Para estos últimos grupos el efecto es significativo, pero en algunos casos sustancialmente menor, en torno al 3 %[9].

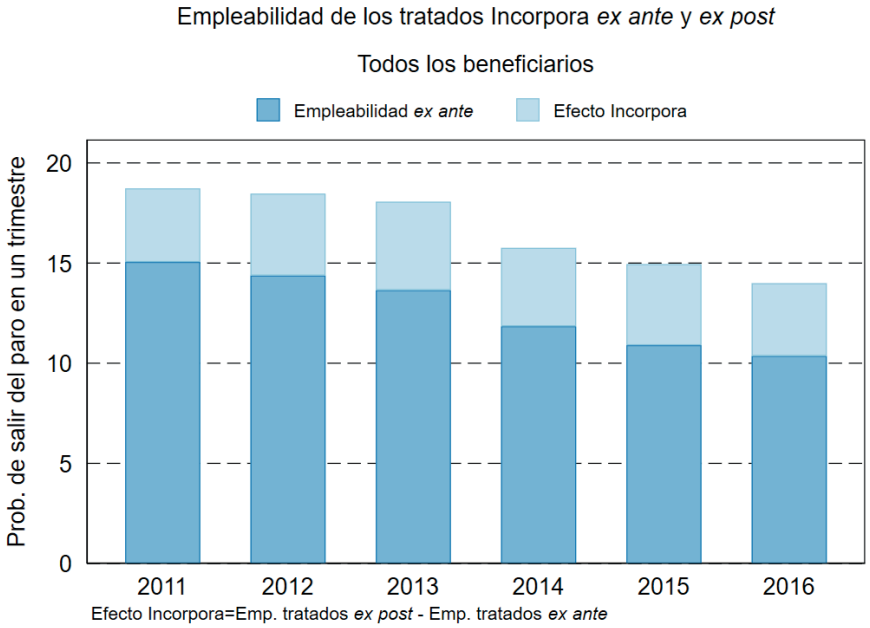
Para el conjunto de todos los beneficiarios encontramos, en un contexto donde la empleabilidad media de los que integran el colectivo cae en el tiempo, un efecto sobre la probabilidad de encontrar empleo significativo y bastante estable en el tiempo. En concreto, la empleabilidad aumenta entre 3 y 4 puntos porcentuales, dependiendo del año (Figura 8). En términos relativos, el efecto es muy importante, entre el 20 y 30 % de la probabilidad promedio *ex ante* de los controles dependiendo del año (Figura 8). El efecto sobre la base de cotización mensual es poco significativo y, si acaso, pequeño al final del periodo considerado (Figura 9). Sí es importante, en cambio, el efecto sobre el número de meses acumulados en el empleo, que se sitúa, dependiendo de la cohorte, entre 3 y 4 meses al cabo de tres años (Figura 9).

Las diferencias entre beneficiarios, en términos de empleabilidad previa, pueden ser un factor importante, ya que es posible que Incorpora tenga un mayor efecto sobre personas ya de por sí más aventajadas en el mercado laboral.

Para comprobar si este es el caso, volvemos a repetir todos estos ejercicios distinguiendo entre tres niveles de empleabilidad previa a la entrada del programa tanto para tratados como para controles. Los resultados de estos nuevos ejercicios (disponibles bajo petición) indican que no existen diferencias en el impacto del programa para aquellos que se dan de alta en Incorpora teniendo un nivel de empleabilidad alto y aquellos que tienen uno bajo. Este efecto es además consistente entre los cinco colectivos analizados.

[9] Estos resultados, así como los que más adelante se citan en función del nivel previo de empleabilidad, están disponibles, a petición del lector interesado, en la versión completa del informe que da origen a este artículo.

FIGURA 8
SALIDA DEL PARO EN UN TRIMESTRE



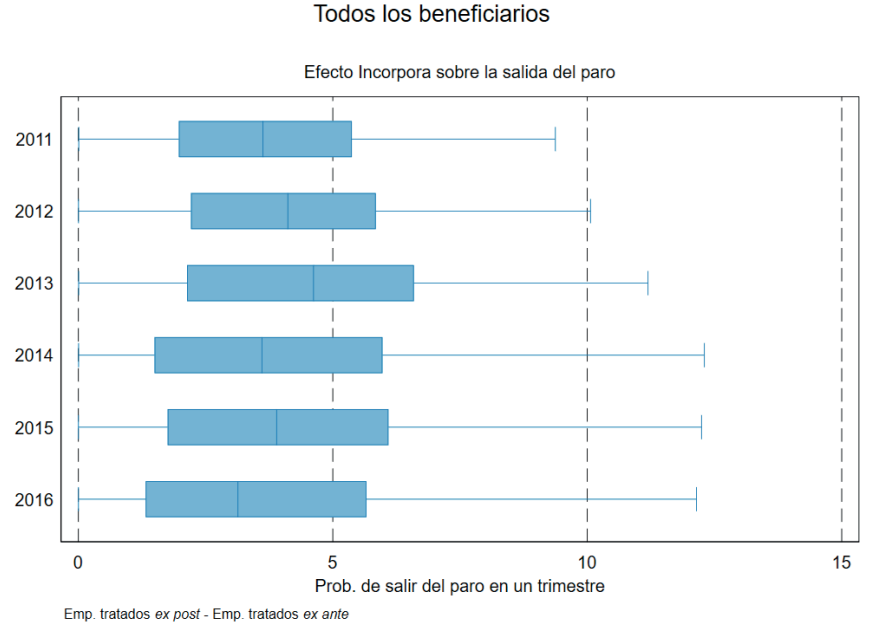
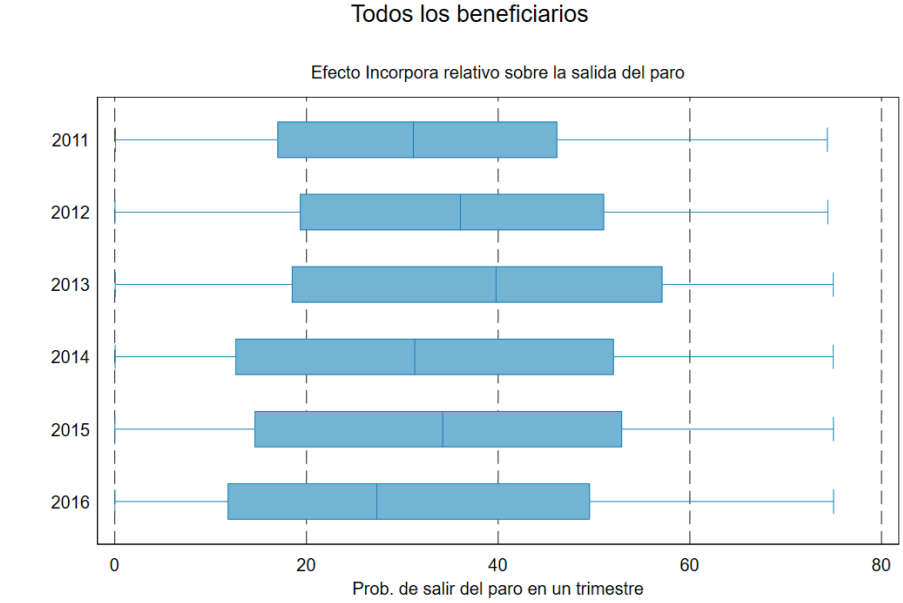


FIGURA 8 *(Continuación)*
SALIDA DEL PARO EN UN TRIMESTRE



FUENTE: Elaboración propia.

FIGURA 9
SALARIOS Y NÚMERO DE MESES ACUMULADOS EN EL EMPLEO

a) Efecto sobre la base de cotización mensual

a*1) *Efecto Incorpora sobre la base de cotización mensual

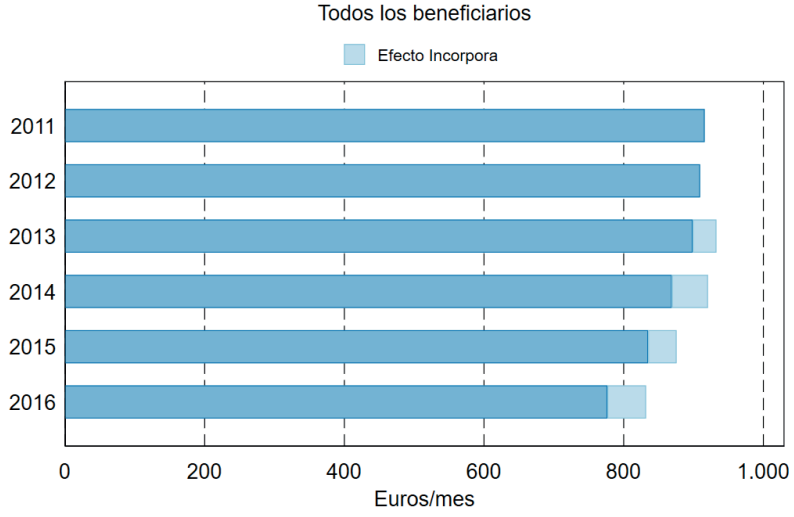
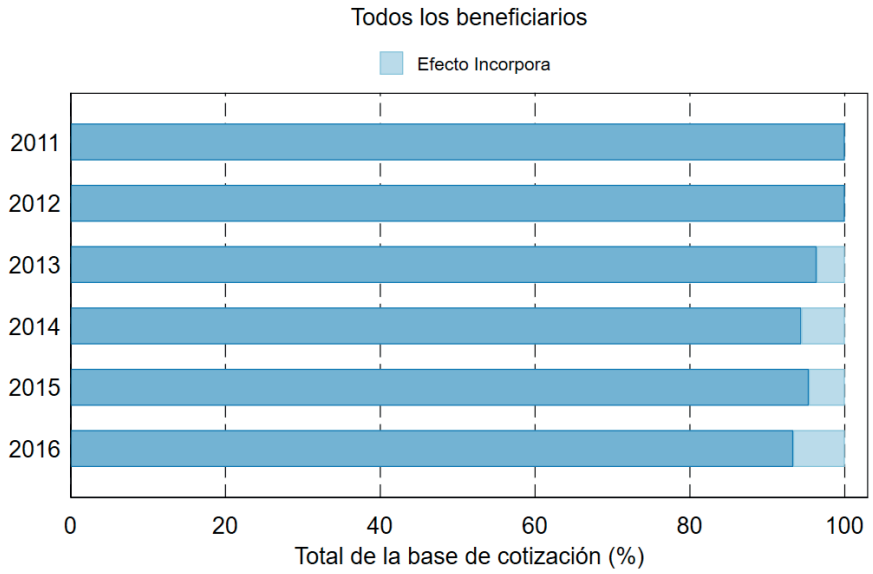
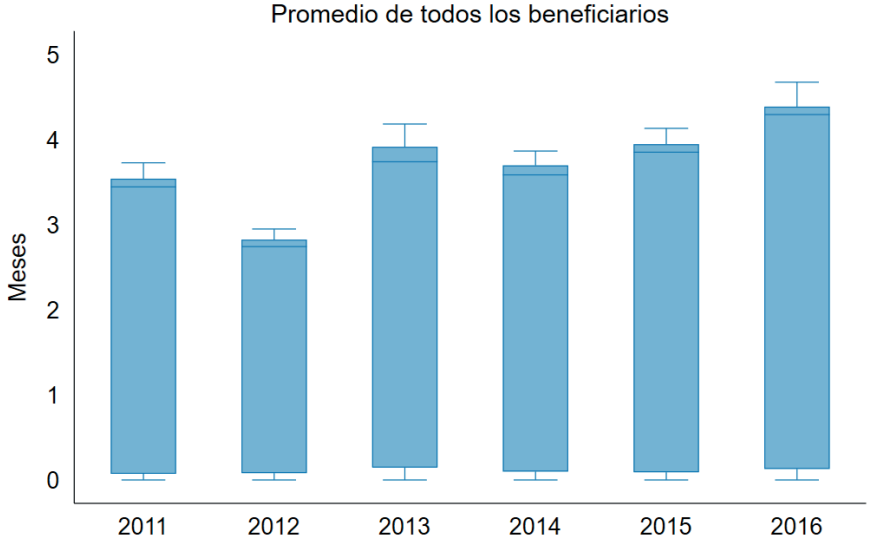


FIGURA 9 ***(Continuación)***
SALARIOS Y NÚMERO DE MESES ACUMULADOS EN EL EMPLEO

a2) Peso relativo del efecto Incorpora sobre la base de cotización mensual



b) Meses acumulados en el empleo



NOTA: Los datos son mes de diferencia respecto a los controles.
FUENTE: Elaboración propia.

5. Conclusiones y recomendaciones

El presente trabajo tiene como objetivo arrojar luz sobre el impacto del programa Incorpora en términos, aunque no exclusivamente, de empleabilidad. Distinguimos dos partes principales: una evaluación interna, de impacto a corto plazo sobre la empleabilidad, a partir de la información suministrada por los técnicos; y otra, complementaria, evaluando el impacto a medio plazo, en base a la información que proporciona el historial laboral inmediatamente posterior a la participación en el programa, de los participantes en el mismo.

Los resultados indican que Incorpora mejora la empleabilidad de aquellos que acceden al programa. Al mismo tiempo, encontramos que la formación contribuye a una mejora de la empleabilidad, pero que este efecto decae en el tiempo.

No obstante, observamos que la asignación de los cursos no parece destinarse a los beneficiarios más vulnerables dentro de cada uno de los colectivos estudiados. Si bien los beneficiarios con una baja empleabilidad que perciben formación representan un porcentaje relativamente alto, también ocurre con aquellos cuyas probabilidades de salir del paro al momento de acceder a Incorpora son relativamente buenas.

La herramienta de perfilado estadístico usada en la primera parte del artículo, adaptada convenientemente para este propósito, podría contribuir a mejorar esta asignación y enfocar el programa en aquellos beneficiarios que más la necesitan. Una forma de mejorar la asignación es generando una evaluación inicial del beneficiario en el momento del alta, donde se indiquen las competencias a mejorar y las herramientas disponibles que mejores resultados generarían.

Referencias bibliográficas

Arellano, F. A. (2010). Do training programmes get the unemployed back to work? A look at the Spanish experience. *Revista de Economía Aplicada, 18*(53), 39-65.

Card, D., Kluve, J., & Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association, 16*(3), 894-931.

Desiere, S., Langenbucher, K., & Struyven, L. (2019). *Statistical Profiling in Public Employment Services: An international comparison*. OECD Social, Employment and Migration (Working Paper No. 224).

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, *14*(1), 153-158.

Felgueroso, F., García-Pérez, J. I., & Jiménez-Martín, S. *Perfilado estadístico: un método para diseñar políticas activas de empleo*. FEDEA y Fundación Ramón Areces.

García-Pérez, J. I. (2008). La muestra continua de vidas laborales: una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada*, *16*(1), 5-28.

Goodman-Bacon, A. (2018). *Difference-in-difference with variation in treatment timing*. NBER, National Bureau of Economic Research (Working Paper Series No. 25018).

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161. https://doi.org/10.2307/1912352

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician, 39*(1), 33-38.

CUADERNOS ECONÓMICOS ICE INFORMACIÓN COMERCIAL ESPAÑOLA

Medidas educativas de refuerzo en lectoescritura: el caso del programa de Ayuda a la Dislexia en la Comunidad de Madrid

Educational reinforcement measures in literacy: the case of the Dyslexia Aid program in the Comunidad de Madrid

Almudena Sevilla*
Pilar Cuevas-Ruiz
University College London

Luz Rello**
Instituto de Empresa

Ismael Sanz
Universidad Rey Juan Carlos

Resumen

En este artículo evaluamos el programa de Ayuda a la Dislexia de la Comunidad de Madrid en el curso 2018-2019 dirigido a alumnos de primaria con dificultades lectoescritoras. El programa consiste en dos herramientas: a) una prueba para comprobar los alumnos que presentan dificultades de lectoescritura, y b) hasta 42.000 ejercicios en una plataforma digital para trabajar y reforzar esas competencias. Un 7,55 % (n=1,022) de los alumnos presentaban dificultades lectoescritoras. Los resultados del análisis muestran que no existe correlación entre la participación del centro en el programa y el rendimiento de los alumnos de esos colegios en las pruebas externas y estandarizadas de Matemáticas y Lengua, pero sí para el caso de Inglés. Al desglosar los resultados por género, se observa una asociación fuerte y positiva del proyecto Ayuda a la Dislexia para las chicas en Lengua e Inglés (y, como se podría esperar, no tanto en Matemáticas), mientras que para los chicos solo es significativa en Inglés. No obstante, se ha comprobado que los centros que participan en esta intervención tenían mejores resultados que los centros no tratados y que sus alumnos provienen de un entorno socioeconómico más elevado. Las diferentes características de los colegios tratados podrían explicar su participación, que era voluntaria, en el programa de Ayuda a la Dislexia. La línea de investigación futura plantea explotar a través de un análisis de diferencias en diferencias que ha habido una nueva convocatoria del programa en el curso 2021-2022, en la que se han incorporado nuevos centros educativos para analizar si existe causalidad.

Palabras clave: *refuerzo educativo, dislexia, programas públicos, educación.*

Clasificación JEL: *I21, I24, I28.*

---

* Almudena Sevilla desea agradecer a European Research Council Consolidator Grant (CoG), SH3, ERC-2017-COG por financiar PARENTIME project. Pilar Cuevas-Ruiz desea agradecer la financiación a las Becas Fundación Ramón Areces para Estudios de Postgrado. XXXIII Convocatoria para Ampliación de Estudios en el Extranjero en Ciencias Sociales 2019/2020.

** Luz Rello es la coordinadora y creadora del programa de Ayuda a la Dislexia que se describe en este artículo. Ismael Sanz era director general de la Consejería de Educación de la Comunidad de Madrid cuando esta iniciativa se puso en marcha en 107 centros educativos.

DOI: https://doi.org/10.32796/cice.2021.102.7317
Recibido: Septiembre 2021 · Aceptado: Octubre 2021

Abstract

In this article, we evaluate the Dyslexia Aid program of the Community of Madrid in the 2018-2019 academic year, aimed at primary school students with reading and writing difficulties. The program consists of two tools: a) a test to check students who have literacy difficulties and b) up to 42,000 exercises on a digital platform to work and reinforce these skills. 7.55% (n=1,022) of the students had reading and writing difficulties. The results of the analysis show that there is no correlation between the participation of the school in the program and the performance of the students at these schools in the external and standardized tests of Mathematics and Language, but there is in the case of English. When disaggregating the results by gender, a strong and positive association of the Dyslexia Help project is observed for girls in Language arts and English (and, as one might expect, not so much in Mathematics), while for boys it is only significant in English. However, schools that participate in this intervention have better results and students from a higher socio-economic background. The different characteristics of the schools could explain their participation, which was voluntary, in the Dyslexia Aid program. The future line of research proposes to exploit through a differences-in-differences analysis that there has been a new call for the program in the 2021-2022 academic year, in which new schools have been incorporated.

Keywords: *remedial education, dyslexia, public programmes, education.*

1. Introducción

Las medidas de apoyo educativo a los alumnos rezagados en educación primaria o secundaria están bajo escrutinio en la mayoría de los países de la OCDE. En Estados Unidos, la atención se centra en la eficiencia, el coste de la inversión y el momento de llevar a cabo las iniciativas de refuerzo a los estudiantes; mientras que en Inglaterra, se debate sobre los posibles itinerarios para los alumnos que no obtienen al menos un grado C en Lengua o Matemáticas en los exámenes estandarizados de secundaria (Lavy *et al*., 2021). En este artículo vamos a realizar un análisis descriptivo de un programa de ayuda a alumnos de primaria con dificultades lectoescritoras en la Comunidad de Madrid. Freyr y Howard-Noveck (2020) señalan que hay poca evidencia de impactos significativos de medidas de refuerzo en el rendimiento en lectura, y que, por el contrario, sí hay más programas de Matemáticas que han demostrado ser eficaces.

En el curso 2018-2019, se llevó a cabo en la Comunidad de Madrid el programa Ayuda a la Dislexia, destinado a colegios, en su mayoría públicos, de primaria. Un total de 107 centros educativos, 100 públicos y 7 concertados participaron en este proyecto ofrecido por la Consejería de Educación y Juventud. El programa, que continúa implementándose en el curso de 2021-2022, está compuesto por unas sesiones de formación y el desarrollo de un *software* informático Dytective que muestra la posibilidad de que un alumno tenga dificultades lectoescritoras, en general, y de dislexia, en particular. Además, la plataforma digital proporciona hasta 42.000 ejercicios para que los alumnos puedan practicar y mejorar esas competencias. En este artículo se pretende realizar un primer análisis descriptivo para comprobar si la participación por parte de los centros educativos madrileños en este programa de Ayuda a la Dislexia se asocia a que sus alumnos muestren mejores resultados en las pruebas externas y estandarizadas de la Comunidad de Madrid en 3.º y 6.º de primaria en Lengua, Matemáticas e Inglés.

Las estimaciones que se presentan no se pueden interpretar como causales, por el conocido sesgo de selección (véase, por ejemplo, Angrist & Pischke, 2014). La participación de los centros educativos en el programa fue voluntaria. Se informó a todos los colegios públicos de la posibilidad de formar parte de este programa de Ayuda a la Dislexia y los centros que mostraron interés pudieron beneficiarse de este *software* de cribado de problemas de competencia lectoescritora y herramienta de tratamiento. Este posible sesgo en la muestra supone una limitación del presente artículo.

A continuación, en el apartado 2, se realiza una revisión de la literatura más reciente sobre programas de refuerzo educativo. En el apartado 3 se describen las principales características del programa de Ayuda a la Dislexia en la Comunidad de Madrid. El apartado 4 presenta las regresiones que muestran correlación entre la participación de los centros educativos y los resultados de sus alumnos en Lengua e Inglés en las pruebas externas y estandarizadas de la LOMCE (Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa). Una contribución relevante del artículo es la desagregación por género de la correlación entre la participación de los colegios en esta iniciativa y el rendimiento de sus estudiantes. El apartado 4 muestra que la asociación es más positiva y fuerte para las chicas que para los chicos, una evidencia sobre la que elaboramos hipótesis basadas en resultados previos de otros estudios. En el apartado 5 se exponen las principales conclusiones.

2. Literatura previa sobre medidas educativas de refuerzo en primaria y secundaria

En los últimos años han surgido evaluaciones de programas asistidos por ordenador, como el de la Comunidad de Madrid que se analiza en este artículo, o libros digitales en educación primaria. Muralidharan *et al*. (2019) analizan un experimento aleatorio de enseñanza asistida por ordenador en cinco centros educativos de enseñanzas medias (entre 4.º de primaria y 3.º de Educación Secundaria Obligatoria, ESO) de barrios de bajos ingresos en Nueva Delhi (India). Una característica clave del *software* empleado es su capacidad para analizar los patrones de errores de los estudiantes y dirigirse con precisión al contenido de mayor demanda cognitiva. Los cuellos de botella conceptuales en los que cada alumno se estanca pueden ser difíciles de diagnosticar para los docentes en una clase con un número elevado de estudiantes. Los 619 participantes del experimento fueron separados aleatoriamente entre un grupo de tratamiento de 314 alumnos a los que se dio la oportunidad de beneficiarse del programa de enseñanza asistida por ordenador y un grupo de control con 305 estudiantes. Los grupos de tratamiento y control no diferían en ninguna dimensión observable. La tasa de asistencia al proyecto entre el grupo de tratamiento fue del 58 %. Muralidharan *et al*. (2019) encuentran que el efecto causal del programa es elevado: 59 % de la desviación estándar en Matemáticas y del 36 % de la desviación estándar en Lengua (Hindi). Estos autores encuentran, además, que el progreso relativo fue mucho mayor para los alumnos que estaban previamente más rezagados, dado que los estudiantes del grupo de control con peores resultados

prácticamente no mejoraron en el tiempo que duró el experimento. El efecto del programa de enseñanza asistido por ordenador analizado por Muralidharan *et al*. (2019) podría ser en realidad el resultado de la combinación del impacto del *software*, pero también el de la tutoría en grupo, y el del tiempo de instrucción adicional. Pero los autores argumentan que gran parte del avance de los alumnos durante el experimento es por el programa asistido por ordenador, porque fue capaz de atender eficazmente a una variedad de estudiantes con distintos niveles de aprendizaje previos dentro de un mismo grupo, dirigiendo con precisión la instrucción al progreso de cada alumno.

Beg *et al*. (2021) realizan dos experimentos aleatorios en la región de Punjab, Pakistán. En el primer experimento participaron 2.999 alumnos de 59 escuelas de enseñanzas medias, del grado 6 (6.º de primaria en España) al 8 (2.º de ESO). Los centros educativos se dividieron de forma aleatoria en un grupo de tratamiento y un grupo de control, para evitar el sesgo de selección derivado de una posible correlación de las características previas de los tratados y su participación en el proyecto educativo. Este primer proyecto incluyó libros digitales, vídeos cortos de expertos educativos con dilatada experiencia explicando conceptos del currículum de Matemáticas y Ciencias, preguntas asociadas a esos vídeos para comprobar el progreso de los estudiantes, o simulaciones de ideas complejas como la fotosíntesis. El experimento se completaba con formación al profesorado para conocer cómo usar en el aula este material y los dispositivos digitales y cómo combinarlos con las clases presenciales. Los resultados de este proyecto, con un coste muy bajo de 9 dólares por alumno, muestran que los estudiantes mejoraron en un 30 % de la desviación estándar sus resultados en Matemáticas y Ciencias, un 60 % más que el grupo control, después de 4 meses. Además, este primer experimento incrementó también en un 26 % de la desviación estándar los resultados de los alumnos en las pruebas externas y estandarizadas que determinan las posibilidades de continuar estudiando en la región. Beg *et al*. (2021) también encuentran que los mayores progresos en los resultados se dieron en los centros educativos y en alumnos que tenían previamente las puntuaciones más bajas, que son probablemente aquellos que más necesidades de formación y apoyo tenían. Los autores muestran, asimismo, que los aumentos en la puntuación fueron mayores en los centros educativos con los profesores menos experimentados y con menos compañeros de su mismo curso con los que compartir la práctica y experiencia docente. El mecanismo a través del cual se produce este impacto positivo del primer experimento aleatorio es el aprendizaje de los profesores a través de los vídeos y la modificación consiguiente en sus métodos de enseñanza y prácticas docentes.

El segundo experimento aleatorio que analizan Beg *et al*. (2021) arroja efectos negativos. Se trata de un programa realizado en la misma región de Punjab, con un grupo también de tratamiento y de control en el que participaron 3.614 estudiantes de 71 centros educativos de enseñanzas medias. El experimento tiene un contenido muy similar al primero descrito previamente, pero los vídeos se incorporan en *tablets* individualizadas para los alumnos, lo que aumenta el coste hasta los 131 dólares por estudiante. Este segundo proyecto no buscaba cambiar la práctica docente de los profesores ni mejorar su formación para que combinasen el nuevo material

con sus clases presenciales. Por el contrario, el segundo experimento se centraba en ofrecer todo el material curricular digital a los alumnos para su seguimiento y aprendizaje tanto en clase como desde sus hogares. El resultado final es que los alumnos que aleatoriamente participan en el segundo proyecto reducen sus puntuaciones en un 40 % de la desviación estándar. Beg *et al*. (2021) concluyen que los diferentes efectos de ambos experimentos se deben a que el primero se realiza con los docentes, con su colaboración, complicidad y complementando su labor. Los vídeos y el material digital se proporcionaban solo a los profesores en el primer experimento, para que lo emplearan en clase con una metodología nueva con la que estuvieran implicados y motivados. De hecho, los autores muestran que la asistencia de los alumnos y docentes participantes en el primer experimento aumentó de forma significativa, al igual que, según su propia información, su nivel de esfuerzo. Por el contrario, el segundo experimento proporcionaba libros digitales y vídeos explicativos de conceptos por parte de expertos educativos, «puenteando» a los docentes.

Otros artículos en la literatura de economía de la educación han evaluado intervenciones centradas en la comprensión lectora, un objetivo que también persigue el programa de Ayuda a la Dislexia en Madrid. Fryer y Howard-Noveck (2020) realizaron un experimento aleatorizado con aproximadamente 1.700 estudiantes de 60 centros públicos de enseñanzas medias en la ciudad de Nueva York. El tratamiento consistía en 2,5 horas de programación diaria en horario extraescolar, que incluía de 45 a 60 minutos de lectura con tutorías de 4 alumnos por tutor para un subconjunto de estudiantes. En total, las escuelas ofrecieron al menos 130 horas de tutoría en una lectura guiada por cada uno de los tres años que duró el proyecto. Un equipo de docentes y expertos en educación secundaria desarrolló un plan de estudios de tutoría detallado (con aproximadamente 50 páginas de material curricular por cada uno de los 150 libros) centrado en lecturas de alto interés y apropiadas para los alumnos del proyecto. Llevando a cabo un análisis de efectos de *intention to treat* (ITT) con errores estándar agrupados a nivel de centro, estos autores muestran que el programa se tradujo en aumentos positivos y significativos en la asistencia de los estudiantes a los centros de 1,2 puntos porcentuales, cuando la media del grupo de control era de 91,9 %. El experimento también proporcionó mejoras, aunque no significativas, en los resultados de las pruebas del Estado en Lengua y un impacto negativo, y también no significativo, en Matemáticas. No hay diferencias en los efectos por nivel de renta de la familia del alumno, dominio del inglés, idioma hablado en casa o características del vecindario. La tutoría sí que pudo haber tenido un mayor impacto para estudiantes de menor capacidad lectora previa. También hay una heterogeneidad importante en los efectos según la raza del alumno. El efecto del tratamiento en los estudiantes de color negro fue de un significativo aumento en 2,0 puntos porcentuales de la asistencia a clase (de 92,4 puntos en el grupo de control) y un incremento significativo en Lengua (del 9 % de la desviación estándar), pero no en Matemáticas. Por su parte, el impacto del experimento en los alumnos hispanos no fue significativo ni en Lengua ni en Matemáticas. Fryer y Howard-Noveck (2020) sostienen que gran parte del diferente impacto del programa entre los alumnos de color y los hispanos se explica por las características promedio de

los tutores con mayor presencia de los primeros: más nivel educativo y de inglés de los tutores, mayor puntuación en el proceso de selección, más experiencia previa en centros educativos e impartiendo tutorías. El principal mecanismo a través del cual, según estos autores, se produce el efecto del programa es que las tutorías individualizan la enseñanza y se proporciona más información *(signaling)* a los alumnos de su progreso.

Gao *et al*. (2018) evalúan la efectividad de un programa de lectura en la mejora de la competencia lectora y en el rendimiento en Lengua y Matemáticas de los alumnos de primaria en la zona rural de la provincia de Guizhou (China). El grupo de tratamiento está compuesto por 2.533 alumnos de 79 clases de 15 escuelas elegidas aleatoriamente entre el condado en donde se llevó a cabo el programa. Por su parte, el grupo de control lo forman 1.575 estudiantes de 49 clases de 15 escuelas elegidas aleatoriamente de entre dos condados vecinos en donde no se implementó este proyecto. Las escuelas son similares en las características observables antes de comenzar el programa de lectura. El grupo de tratamiento participó en uno de los siguientes programas: *a*) la dotación a los colegios de libros extraescolares de cuentos, dibujos y ciencias naturales; *b*) esta dotación unida a una formación de las autoridades educativas a los docentes del colegio sobre la forma de instruir la asignatura de Lengua y el análisis de pasajes clave de lecturas, desarrollo de vocabulario y resumen del contenido; y *c*) dotación de libros para el colegio y una formación a los docentes realizada por una ONG centrada en cómo enseñar a los alumnos a leer de forma independiente y a promover actitudes positivas hacia la lectura, tanto entre los alumnos como entre los profesores. Pues bien, Gao *et al*. (2018) solo encuentran evidencia de efectos positivos del último de los programas, con un impacto muy elevado en relación con los efectos encontrados habitualmente en otros estudios. En concreto, los alumnos de los colegios que recibieron libros y formación a sus docentes de la ONG tienen un 61 % de la desviación estándar más de resultados en comprensión lectora que los del grupo de control. También hallan un impacto del 35 % de la desviación estándar en las pruebas estandarizadas posteriores de Matemáticas y un 29 % en las de Lengua. El efecto en Matemáticas se basa en que la mejora de la competencia lectora aumenta el vocabulario y la capacidad de comprensión de los problemas de matemáticas e incluso el pensamiento crítico. Los efectos del programa C son significativos mientras que los del programa B y A no son distintos de cero. Los autores creen que el impacto es diferente por la formación de los docentes de la ONG, que trabaja en métodos, actividades y materiales adaptados a cada edad para inculcar el gusto por la lectura y que esta se desarrolle de forma independiente por parte de los alumnos, mientras que la formación a los docentes impartida por las autoridades se centra más en las competencias lectoras, que van a ser después evaluadas en las pruebas estandarizadas. De hecho, Gao *et al*. (2018) muestran que los docentes que fueron formados por la ONG llevaron a cabo más actividades de lectura después del programa en clase, crearon más grupos para debatir sobre el contenido de los libros y animaron a los alumnos a que hicieran presentaciones, de modo que los estudiantes pudieran estar más implicados con sus lecturas.

Guryan *et al*. (2016) también encuentran que la entrega de libros por sí sola no tiene efectos positivos en la comprensión lectora de los alumnos. Estos autores

analizan un experimento aleatorio controlado en el que 415 alumnos de 4.º y 5.º de primaria de un distrito urbano del noreste de Estados Unidos reciben un libro a la semana durante diez semanas de verano (proyecto READS) y seis lecciones de estrategias para promover la lectura independiente, mientras que el grupo de control no participa en la intervención hasta la primavera del año siguiente. El grupo de tratamiento se divide entre un subgrupo de alumnos que solo obtienen los libros y otro subgrupo que, además, tiene un incentivo en forma de puntos por cada lectura que les da acceso a recoger regalos. Los participantes en el proyecto READS, en ambas modalidades, tenían que hacer una ficha por cada libro contando la parte que más les había gustado. Los autores hallan efectos positivos en la comprensión lectora solo cuando el nivel de la lectura es el apropiado para las competencias previas del estudiante y en aquellos alumnos que antes del experimento ya tenían mayor motivación por leer libros. Por el contrario, no se encuentra ningún impacto en aquellos menos motivados y que eran los destinarios prioritarios del proyecto. Guryan *et al*. (2016) señalan que los estudiantes con menos motivación intrínseca que, quizá porque descuentan más el futuro o tienen menos capacidad de control, se implican menos en las tareas escolares en general, y en la lectura en particular, responden también menos a la motivación extrínseca que ofrecía uno de los tratamientos. Para los alumnos con mayor motivación intrínseca, la existencia de incentivos a la lectura en forma de regalos sí tiene efectos positivos y significativos tanto en una prueba estandarizada de comprensión lectora en otoño de ese mismo año como en test general de primavera del siguiente. En concreto, una desviación estándar por encima en motivación intrínseca (medido antes de la intervención por el Motivation Reading Questionnaire) conduce a un incremento en un 29,5 % de la desviación estándar en comprensión lectora y un 20,4 % también de la desviación estándar en la prueba general de Massachussets.

La falta de formación del profesorado o de incentivos a los alumnos puede explicar que Goux *et al*. (2017) no encuentren efectos positivos para un programa similar para estudiantes de 1.º de primaria en colegios de París y Lille, en Francia. Se trata de un proyecto cuyo objetivo es desarrollar el gusto por la lectura y la competencia lectora a alumnos que viven en hogares desfavorecidos. El experimento aleatorio controlado se llevó a cabo en 109 colegios, cerca de dos tercios de tratamiento y el resto de control, de áreas deprimidas. El programa fue implementado también por una ONG y consistía en seis horas a la semana extraescolares de apoyo durante casi seis meses, dedicado a la expresión oral, juegos con palabras y textos, y la lectura de cuentos tanto por parte del monitor como de los alumnos. Cada profesor seleccionaba cinco alumnos que no tenían un contexto en el hogar que favoreciera la lectura y con riesgo de quedar rezagados del resto de los compañeros. Goux *et al*. (2017) encuentran que el programa mejora en un 17 % de la desviación estándar el gusto por la lectura en los alumnos tratados, e incluso también por las Matemáticas. Sin embargo, este aumento del disfrute de la lectura no se traslada a las competencias en lectura ni en Matemáticas de los alumnos que no experimentan ningún incremento significativo. Tampoco hay externalidades positivas del programa en el gusto por la lectura de los alumnos que no participan en el programa, pero están en la misma clase que los 5 alumnos que sí intervienen en el

proyecto. Goux *et al*. (2017) argumentan que el hecho de que el programa sustituyera a otras actividades de refuerzo y apoyo también extraescolar, que suelen implementar las escuelas francesas situadas en zonas desfavorecidas, puede explicar que no se encuentre un impacto significativo en las competencias en lectura ni en Matemáticas.

Machin *et al*. (2018) evalúan también un programa de enseñanza de lectura en primaria, en este caso en Inglaterra, a través de un análisis de diferencias-en-diferencias. No se trata, como en los anteriores artículos, de un experimento aleatorio. La intervención, con un coste económico relativamente bajo, modificó la pedagogía en la enseñanza en lectura a través de cursos de formación a los docentes. Se llevó a cabo en dos fases, una primera piloto llevada a cabo en el curso 2005-2006 en 18 autoridades locales educativas, y su ampliación a 50 autoridades locales del país a partir del 2006-2007. En 2008-2009 y 2009-2010 entraron centros educativos de otras 50 autoridades locales por cada curso. El enfoque pedagógico se cambió desde el «lenguaje total» en el que se enseña el lenguaje directamente a través de cuentos, dibujos y libros a la «fonética sintética», una técnica en la que se enseñan primero los sonidos de las letras y las combinaciones de letras y después el método para formar palabras completas. Los autores aprovechan que la intervención se implementó en años sucesivos para comparar los resultados de los centros educativos de las 18 autoridades locales que participaron en la fase piloto de 2005-2006 (grupo de tratamiento 1) y los 50 del curso 2006-2007 (grupo de tratamiento 2) con los que lo empezaron en 2007-2008 y 2008-2009 (grupo de control). Machin *et al*. (2018) encuentran efectos positivos en la comprensión lectora de los alumnos de 5 años de hasta un 30 % de la desviación estándar en la fase piloto y el 22 % en segunda fase. A los 7 años, dos tercios de ese impacto habían desaparecido ya, aunque el impacto continuaba siendo positivo y significativo. Cuando esa misma cohorte tenía 11 años, no había ningún efecto ya significativo para el conjunto de los alumnos. Sin embargo, para los estudiantes cuya lengua nativa no es el inglés y los que son elegibles para las becas de comedor, el impacto, de más de un 10 % de la desviación estándar y significativo, continúa presente a los 11 años. Es decir, que el programa reduce la diferencia de los estudiantes que provienen de entornos desfavorecidos o para los que el inglés no es su lengua nativa, con respecto al conjunto de la clase. Machin *et al*. (2018) también encuentran efectos positivos del programa de «fonética sintética» en Lengua y Matemáticas, aunque también decaen cuando los alumnos tienen 11 años.

3. Principales características del programa de Ayuda a la Dislexia en la Comunidad de Madrid

La Asociación Internacional de la Dislexia (International Dyslexia Association, IDA, 2002) define a la dislexia como: «Dificultad específica del aprendizaje de origen neurobiológico. Se caracteriza por dificultades en el reconocimiento preciso de las palabras y déficit en la decodificación de la lectoescritura. Estas dificultades resultan de un déficit en el componente fonológico del lenguaje [...]. Como consecuencias secundarias pueden presentarse problemas en la comprensión de la lectura y una experiencia lectora reducida que puede afectar el incremento del vocabulario y de la base de conocimientos».

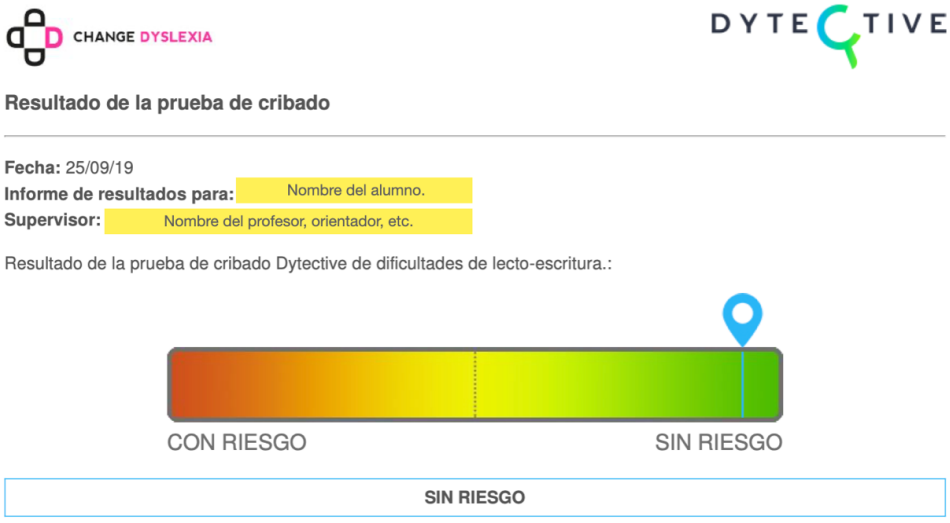
Hasta donde alcanza nuestro conocimiento, en España solo se habían realizado dos estudios principales de prevalencia de dislexia y con resultados dispares. Jiménez *et al*. (2009) para las Islas Canarias con una muestra de 50.000 alumnos de educación primaria (entre 2.º y 6.º curso de primaria) indican una prevalencia del 8,6 % de alumnos con dificultades específicas de aprendizaje, siendo la estimación de dislexia de un 3,2 %. Carrillo-Gallego *et al*. (2011) llevan a cabo un estudio en Murcia con una muestra de 2.012 alumnos de enseñanza primaria y estiman una prevalencia del 15,2 % para 2.º curso de primaria, 9,3 % para 4.º curso y 10,9 % para 6.º de primaria. La detención de la dislexia es compleja y más especialmente en lenguas más transparentes como el español (es decir, lenguas cuya correspondencia entre grafema –letra– y fonema –sonido– es más regular), en las que los niños pueden camuflar sus dificultades en los primeros años de primaria.

Pues bien, la Consejería de Educación y Juventud de la Comunidad de Madrid implementó un programa de Ayuda a la Dislexia que incluía un test de cribado de posibles problemas de lectoescritura y dislexia entre alumnos de primaria. Un total de 107 centros educativos, 100 públicos y 7 concertados en la Comunidad de Madrid en el curso 2018-2019 participaron en el proyecto. Los 100 colegios públicos, en los que se concentra nuestro análisis, representan el 12,5 % de los 802 Colegios de Educación Infantil y Primaria (CEIP) públicos que hay en la Comunidad de Madrid.

El programa Ayuda a la Dislexia, tiene como objetivo la detección y corrección de este trastorno de la lectoescritura a través de la plataforma Dytective. Está compuesto por dos partes: *i*) un test de cribado de riesgo de dificultades de lectoescritura (Dytective Test) y *ii*) un videojuego para la estimulación y mejora de las habilidades de lectoescritura o apoyo al tratamiento (Dytective U). Además, este proyecto de refuerzo está acompañado por unas sesiones de formación dirigidas a directores de colegio, profesores de Lengua, orientadores, maestros PT (Pedagogía Terapéutica) y maestros AL (Audición y Lenguaje). La prueba de cribado tiene 32 ítems con una duración aproximada en total de 15 minutos y una sensibilidad del 80,4 % para los estudiantes con dislexia de entre 7 y 17 años. La sensibilidad se define como positivos verdaderos dividido entre positivos verdaderos más falsos negativos. El resultado de esta prueba de cribado es un informe (Figura 1) que indica el riesgo de tener dificultades de lectoescritura. La prueba de cribado integra un algoritmo de aprendizaje automático basado en árboles de decisión (Random Forests) que fue evaluado con 3.644 participantes por ordenador y 1.395 participantes por *tablets* donde el 10,7 % y el 10,6 % presentaron dislexia diagnosticada (Rello *et al*., 2020).

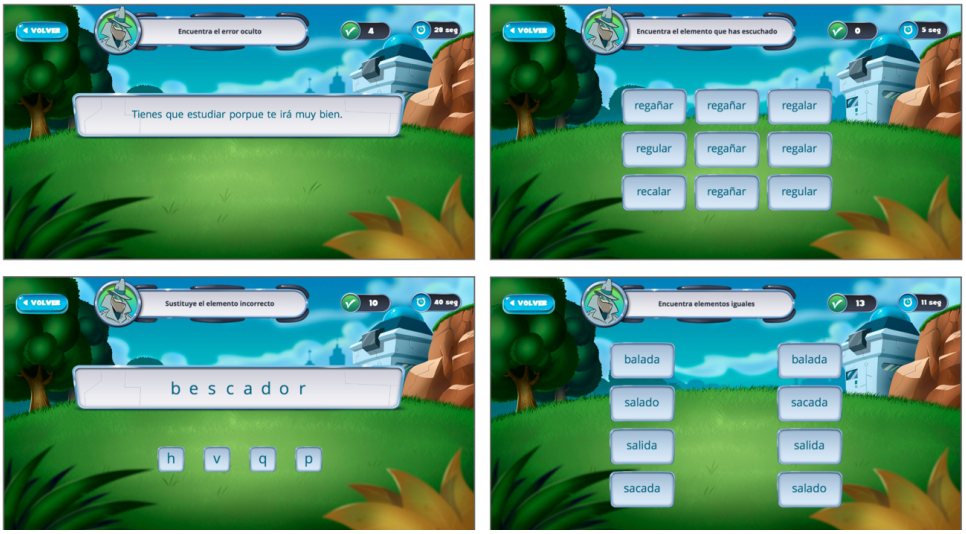
La segunda parte de Dytective (llamada Dytective U) es un videojuego compuesto por 42.000 ejercicios destinados a la estimulación y refuerzo de la capacidad lectoescritora de los alumnos. Los ejercicios lingüísticos están diseñados teniendo en cuenta los diferentes niveles lingüísticos y los patrones encontrados en los errores que cometen los niños con dificultades de aprendizaje (Figura 2). Los ejercicios se secuencian de acuerdo con el rendimiento de la persona, en relación con 24 habilidades cognitivas específicas dirigidas tanto a las debilidades como a las fortalezas cognitivas, creando para cada usuario un modelo único de aprendizaje (Rello, 2018).

FIGURA 1
EJEMPLO DE INFORME DYTECTIVE TEST



FUENTE: Elaboración propia.

FIGURA 2
EJEMPLO DE EJERCICIOS LINGÜÍSTICOS DE DYTECTIVE U

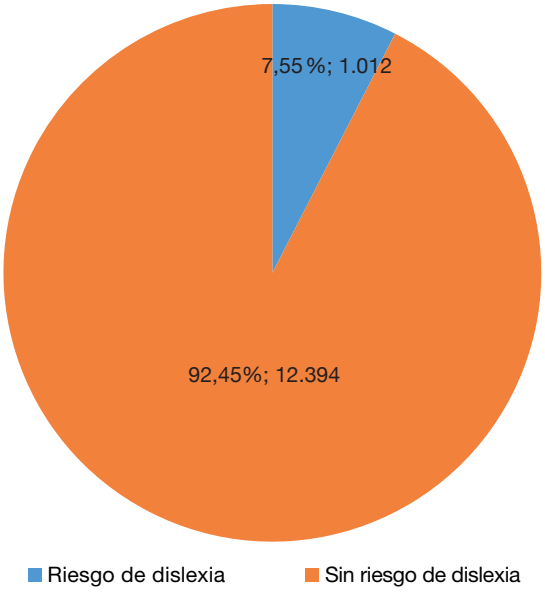


FUENTE: Elaboración propia.

El protocolo aplicado en la Comunidad de Madrid para el uso de Dytective U recomienda que los alumnos efectúen 4 sesiones por semana durante un mínimo de 8 semanas. En total, los alumnos que siguen por completo el protocolo llevan a cabo 8 retos por semana, con el objetivo de realizar un total de 64 retos en 8 semanas. El seguimiento del alumno se puede realizar a través de los informes de la herramienta (Figura A1 del Apéndice).

La Figura 3 muestra el número de alumnos que realizaron el test de cribado de Dytective y obtuvieron un diagnóstico válido en el curso 2018-2019 en la Comunidad de Madrid. En total, 16.830 estudiantes participaron en la prueba de Dytective en esta región, pero únicamente un 79,66 % lo finalizó correctamente (13.406 alumnos). El *software* muestra que 1.012 alumnos podrían tener problemas de lectoescritura, entre ellos dislexia. Dicho número supone el 7,55 % de los estudiantes que realizaron el test de Dytective correctamente, una proporción en línea con la prevalencia que encuentra la investigación internacional y por debajo del 10 % que se maneja en el debate público.

FIGURA 3
PREVALENCIA DE LA DISLEXIA SEGÚN DYTECTIVE TEST



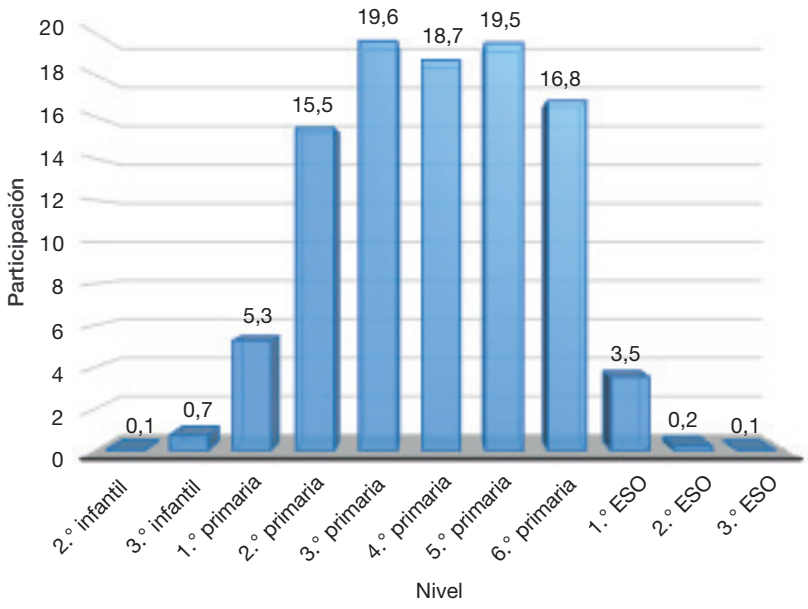
NOTA: Proporción de alumnos con riesgo y sin riesgo de dislexia según el test de cribado en Dytective Test bajo el programa Ayuda a la Dislexia en el curso escolar 2018-2019. Número de alumnos en términos absolutos entre paréntesis. Muestra: 3.424 obtienen resultado nulo; 13.406 reciben un diagnóstico. Estos datos no son necesariamente representativos sobre la prevalencia de la dislexia en los colegios. Si bien en el protocolo se indicó a los colegios que se aplicara el test al total de los alumnos, algunos colegios han podido preseleccionar a los alumnos a los que se les aplica el test.

FUENTE: Elaboración propia.

Si desglosamos los datos por nivel académico, la Figura 4 revela que el 95,3 % de los alumnos que realizaron el test de cribado se encontraban en primaria. Un 3,8 % de los estudiantes estaban matriculados en ESO y el resto en 2.º o 3.º de infantil. El programa se destinaba principalmente a los alumnos de primaria, pero los datos muestran que algunos centros emplearon Dytective para los alumnos de segundo ciclo de infantil, que también se encuentran en los Colegios de Educación Infantil y Primaria (CEIP). También hay estudiantes de la ESO, que en este caso provienen de centros concertados que ofrecen primaria y secundaria y de algunos CEIPSO (Colegios de Educación Infantil, Primaria y Secundaria Obligatoria) públicos que tienen ambos niveles educativos.

El Cuadro 1 muestra la proporción de alumnos con riesgo de dislexia por nivel académico. Es interesante observar que la posible prevalencia de la dislexia alcanza su pico en 1.º y 2.º de primaria, cursos a partir de los cuales comienza a bajar del umbral del 10,0 %. Es un resultado que refleja que los problemas de lectoescritura se concentran en los primeros dos cursos de primaria. El Cuadro 1 es una sección cruzada, una fotografía de los alumnos en el curso 2018-2019 y no se puede interpretar como un panel de seguimiento de los mismos estudiantes a lo largo de diferentes

FIGURA 4
PARTICIPACIÓN EN LA PRUEBA DE CRIBADO POR NIVEL ACADÉMICO
(En %)



NOTA: Proporción de estudiantes que realizaron el test de cribado (Dytective Test) en 2018-2019 por curso académico.

FUENTE: Elaboración propia.

CUADRO 1
ALUMNOS CON RIESGO DE DISLEXIA POR NIVEL ACADÉMICO EN EL CURSO 2018-2019

Nivel académico	Sin riesgo de dislexia	Riesgo de dislexia	Total de alumnos que obtienen un diagnóstico	Alumnos con riesgo de dislexia (%)
2.º infantil	1	0	1	0,00
3.º infantil	5	0	5	0,00
1.º primaria	210	38	248	15,32
2.º primaria	1.591	382	1.973	19,36
3.º primaria	2.576	233	2.809	8,29
4.º primaria	2.587	115	2.702	4,26
5.º primaria	2.624	105	2.729	3,85
6.º primaria	2.324	91	2.415	3,77
1.º ESO	442	44	486	9,05
2.º ESO	29	3	32	9,38
3.º ESO	5	1	6	16,67
Total que han realizado Dytective Test	**12.394**	**1.012**	**13.406**	

NOTAS: Prevalencia de dislexia según participación y cribado en la herramienta Dytective Test en 2018-2019 por curso académico. Porcentaje realizado sobre aquellos alumnos que finalizan el test y obtienen una valoración.
FUENTE: Elaboración propia.

niveles. Aun así, se podría inferir que una parte de los alumnos con un posible riesgo de dislexia en 1.º y 2.º de primaria, en realidad tenían dificultades en lectoescritura que superan reforzando y trabajando estas competencias a partir de 3.º de primaria. De este modo, la prevalencia del 19,36 % en 2.º de primaria se reduce hasta el 8,29 % en 3.º de primaria, para estabilizarse finalmente entorno al 4 % a partir de 4.º de primaria. Esta bajada en la prevalencia de dislexia según aumenta el nivel académico es consistente con la literatura anterior (Carrillo Gallego *et al*., 2011). Los resultados que se alcanzan para la ESO tienen más que ver con el tamaño de la muestra, que es muy pequeño y que se limita a un número pequeño de centros educativos en los que, además, la muestra de alumnos que realizó el test podría estar sesgada hacia aquellos que tienen problemas persistentes de lectoescritura.

El Cuadro 2 muestra que la participación en las pruebas de Dytective Test es equilibrada por género, con una ligera mayor presencia de chicos (51,67 %) que de chicas (48,33 %). En el conjunto de estudiantes en primaria también hay una proporción algo mayor de alumnos que de alumnas. Los datos oficiales del Ministerio de Educación y Formación Profesional señalan que en el curso 2018-2019 había un 48,6 % de chicas en el total del alumnado de la Comunidad de Madrid (Fuente: series del alumnado matriculado de la Estadística de las Enseñanzas no universitarias. https://bit.ly/3ihnfUo).

CUADRO 2
LA REALIZACIÓN DEL TEST DE DISLEXIA POR GÉNERO

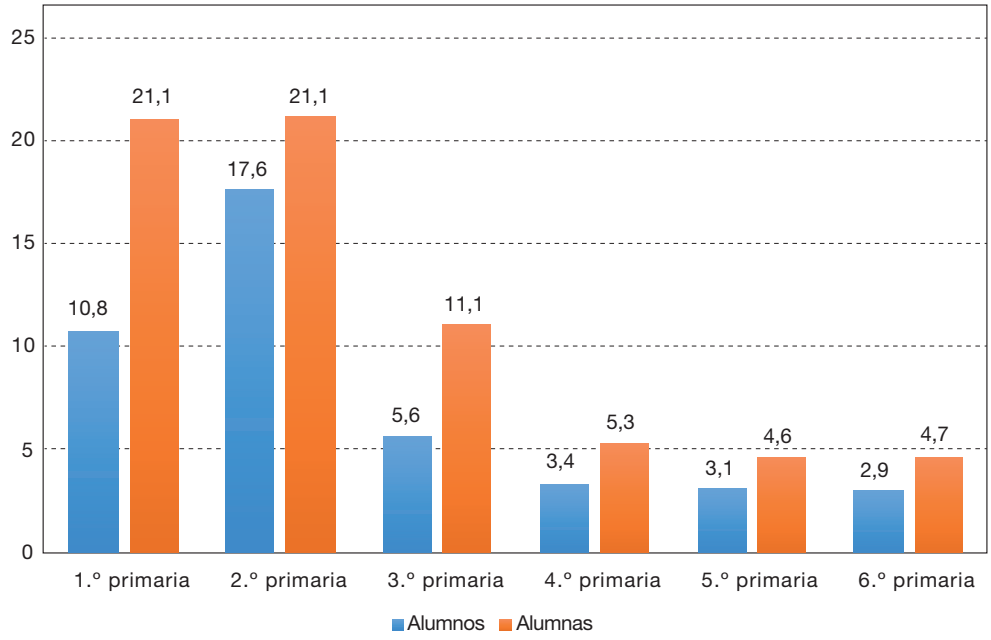
	Total	**% del total**
Alumnas	8.134	48,33
Alumnos	8.696	51,67
Total	**16.830**	**100,00**

NOTA: Proporción de alumnos por género entre 2.º de infantil y 3.º de ESO que han realizado el test de cribado (Dytective Test) en el curso 2018-2019.
FUENTE: Elaboración propia.

A continuación, en la Figura 5, desglosamos los resultados del cribado por género y curso académico para primaria. En el conjunto de la muestra, un 6,02 % de los alumnos chicos son diagnosticados con riesgo de dislexia, mientras que para las chicas alcanza el 9,04 %. Para todos los cursos se da una mayor proporción de alumnas con riesgo de dislexia que alumnos. Estos datos por género contrastan con los datos oficiales del Ministerio de Educación y Formación Profesional. En la Estadística de las Enseñanzas no universitarias, alumnado con necesidad específica de apoyo educativo (https://bit.ly/2UXU0he) se ofrece información para España y por comunidades autónomas de: *a*) los alumnos con necesidades educativas especiales y *b*) otro alumnado con necesidad específica de apoyo educativo. Dentro del apartado *b*) se contemplan 7 categorías, una de las cuales es «trastornos del aprendizaje», que se define como «trastornos específicos del desarrollo del aprendizaje escolar de la lectura, la expresión escrita, el cálculo y otros trastornos no especificados. Se incluyen, entre otros, los siguientes: dislexia, disortografía y discalculia» (Ministerio de Educación y Formación Profesional: Enseñanzas no universitarias. Alumnado con necesidad específica de apoyo educativo). No hay información desagregada solo sobre el número de alumnos con dislexia, sino que los datos están agregados con algunos trastornos del aprendizaje similares. En cualquier caso, esta fuente oficial muestra que en la Comunidad de Madrid hay 3.591 alumnos en educación primaria y 3.676 en ESO con «trastornos del aprendizaje», que incluye dislexia, pero también otros como disortografía o discalculia. Del total de alumnos con trastornos de aprendizaje en la Comunidad de Madrid, un 63,4 % en primaria y un 61,5 % en ESO son chicos. Entre la población general, los niños son diagnosticados más frecuentemente que las niñas. En Dytective Test, la conclusión es la contraria, es decir, se da una mayor prevalencia entre las alumnas.

El hecho de que los datos administrativos muestren una mayor prevalencia de los trastornos del aprendizaje entre los chicos, mientras que en Dytective Test es superior para las chicas, se podría deber a que para los sistemas educativos es más difícil de detectar a las alumnas con dislexia que a los alumnos. Las razones que se han buscado para explicar este hecho son diversas siendo la más frecuente el más bajo rendimiento en lectura de los chicos, de modo que las chicas pasan más

FIGURA 5
PROPORCIÓN DE ALUMNOS CON RIESGO DE DISLEXIA POR GÉNERO Y NIVEL EN EL CURSO 2018-2019 (En %)

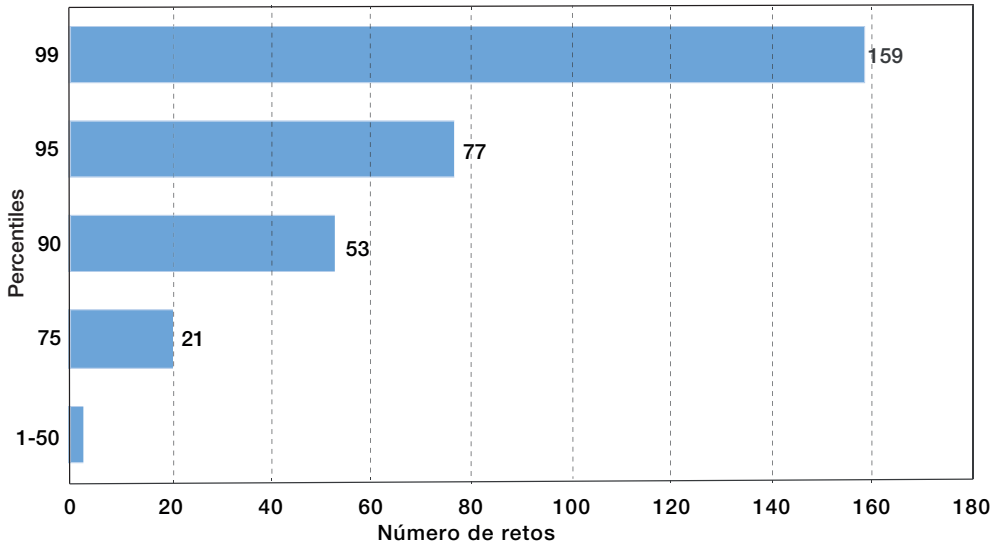


NOTAS: Proporción de alumnos y alumnas con riesgo de dislexia según el Dytective Test sobre el total que ha obtenido una valoración por curso académico.
FUENTE: Elaboración propia.

desapercibidas (Arnett *et al*., 2017). Quizás por este hecho el Dytective test ha detectado a más niñas que a niños, que estarían posiblemente ya diagnosticados. Cabe notar que el cribado Dytective test detecta por igual tanto a niños como a niñas (Rello *et al*., 2020) y que la dislexia no tiene diferencias de género. Varios estudios de prevalencia de dislexia no han encontrado diferencias significativas entre niños y niñas tanto para el inglés (Shaywitz & Shaywitz, 2003) como en español (Jiménez *et al*., 2011).

La información a partir de la Figura 6 siguiente se basa en la segunda vertiente del programa, la de los ejercicios de la herramienta Dytective U. En concreto, se muestra el número de retos llevados a cabo por aquellos alumnos que han sido cribados con riesgo de dislexia por Dytective Test. La mitad de los detectados con riesgo de dislexia realizaron únicamente 3 retos o menos. Sin embargo, un 25 % de ellos realizó más de 21 retos, superando el protocolo completo de 68 retos cerca de un 10 %. Es decir, una parte importante de los alumnos a los que iba dirigido el programa de Ayuda a la Dislexia no realizó, o lo hizo con muy poca intensidad, los retos para trabajar y reforzar las capacidades lectoescritoras.

FIGURA 6
DISTRIBUCIÓN DEL NÚMERO DE RETOS EN DYTECTIVE U DE ALUMNOS CON RIESGO DE DISLEXIA
(En %)



NOTAS: Distribución del número de retos realizados a través de la herramienta Dytective U en el curso 2018-2019 de alumnos que han sido detectados con riesgo de dislexia por Dytective Test matriculados entre 2.º de infantil y 3.º de ESO.
FUENTE: Elaboración propia.

El Cuadro 3 muestra las características de los centros públicos que participaron en el programa de Ayuda a la Dislexia en comparación con los que no lo hicieron. La participación de los centros educativos en el programa fue voluntaria. Se informó a todos los colegios públicos de la posibilidad de formar parte de este proyecto y los centros que mostraron interés pudieron beneficiarse de este *software* de cribado de problemas de lectoescritura y de herramientas de tratamiento para reforzar esta competencia. Por tanto, los centros que han formado parte del proyecto podrían tener características propias que explicaran las posibles diferencias en los resultados en las pruebas externas y estandarizadas con respecto a los colegios que no mostraron interés en utilizar Dytecitive Test y Dytective U. Características como una posible mayor sensibilidad hacia los alumnos con dificultades en lectoescritura, una más elevada predisposición a integrarse en nuevas iniciativas educativas que llevan más tiempo de trabajo y compromiso, y una vocación para estar atentos y responder a las comunicaciones de la Consejería de Educación podrían hacer más proclives a los centros a participar en el programa, al mismo tiempo que conducen a mejorar los resultados académicos de sus alumnos. De hecho, el Cuadro 3 muestra que los

CUADRO 3

DESCRIPTIVOS DE CENTROS TRATADOS Y NO TRATADOS POR EL PROGRAMA AYUDA A LA DISLEXIA EN EL CURSO ACADÉMICO 2018-2019

	Centros incluidos en el programa Ayuda a la Dislexia				Centros no incluidos en el programa Ayuda a la Dislexia				Definición
	Media	Desv. est.	Mín.	Máx.	Media	Desv. est.	Mín.	Máx.	
Nota en Lengua	503,350	92,87	165	930	500,551	96,40	40	930	Variable continua indicativa del rendimiento en Lengua Castellana.
Nota en Inglés	505,170	93,30	145	846	498,423	93,74	99	846	Variable continua indicativa del rendimiento en Inglés.
Nota en Matemáticas	495,283	93,64	213	898	492,517	93,14	20	898	Variable continua indicativa del rendimiento en Matemáticas.
Mujer	0,481	0,50	0	1	0,483	0,50	0	1	Variable binaria que toma valor 1 si el sexo es femenino y 0 si es masculino.
% de alumnos/as inmigrantes en el centro	1,009	1,89	0	22	1,133	1,86	0	26	Variable continua indicativa del porcentaje de alumnos que han nacido fuera de España.
Número de libros en el hogar	3,260	1,22	1	5	3,195	1,23	1	5	Variable categórica que toma valor 1 si el alumno/a reporta menos de 10 libros; 2 de 11 a 50 libros; 3 de 51 a 100 libros; 4 de 101 a 200 libros; y 5 de más de 200 libros.
Educación de la madre	5,126	1,77	1	9	5,030	1,79	1	9	Variable categórica que toma valor 1 si el nivel educativo de la madre del alumno/a es inferior a primaria; 2 si es menor a EGB o ESO; 3 si posee EGB o ESO; 4 si Bachillerato, BUP, COU, FPI; 5 si Formación Profesional superior (FPII); 6 si Diplomatura Universitaria; 7 si Licenciatura o Grado; 8 si Máster Universitario; y 9 si posee Doctorado.
Educación del padre	4,737	1,79	1	9	4,678	1,80	1	9	Variable categórica que toma valor 1 si el nivel educativo del padre del alumno/a es inferior a primaria; 2 si es menor a EGB o ESO; 3 si posee EGB o ESO; 4 si Bachillerato, BUP, COU, FPI; 5 si Formación Profesional superior (FPII); 6 si Diplomatura Universitaria; 7 si Licenciatura o Grado; 8 si Máster Universitario; y 9 si posee Doctorado.

NOTA: Muestra: Alumnos de 3.º y 6.º de primaria evaluados en las pruebas externas de la Comunidad de Madrid en el curso 2018-2019 en colegios públicos.
FUENTE: Elaboración propia.

centros educativos públicos que participaron en el programa de Ayuda a la Dislexia tienen unos resultados ligeramente más elevados que los no tratados en las pruebas externas y estandarizadas de Matemáticas, Lengua e Inglés de 3.º y 6.º de primaria en el curso 2018-2019. Se trata de exámenes con las mismas preguntas y que se llevan a cabo el mismo día y a la misma hora en todos los colegios públicos, concertados o privados de la Comunidad de Madrid. Se desarrollaban como consecuencia de la normativa sobre la realización de pruebas externas en 3.º y 6.º de primaria, así como 4.º de la ESO de la LOMCE. Las madres y padres de los colegios que participaron en el proyecto tienen un mayor nivel educativo en promedio y disponen de más libros en sus hogares, al mismo tiempo que tienen una menor proporción de alumnos nacidos fuera de España. Las puntuaciones se computan como en PISA, con media de 500 y desviación típica de 100.

4. Asociación entre el programa de Ayuda a la Dislexia y los resultados en las pruebas externas y estandarizadas de la Comunidad de Madrid

El Cuadro 4 muestra la correlación, que no causalidad, entre la participación del centro educativo en el programa Ayuda a la Dislexia y los resultados de los alumnos. Los coeficientes encontrados en el Cuadro 4 no se pueden interpretar como causales, por el conocido problema del sesgo de selección (véase Angrist & Pischke, 2014). De hecho, el Cuadro 3 ya muestra que los colegios tratados tienen mejores resultados en Matemáticas, Lengua, y, sobre todo, Inglés, y que el nivel socioeconómico de las familias es también más elevado. Este posible sesgo de selección en la muestra es una limitación del presente informe. Entre los factores asociados a los resultados de los alumnos en las pruebas externas y estandarizadas de 3.º y 6.º de primaria se incluye, junto a variables individuales del propio alumno, una *dummy* que refleja si el centro en el que estudia ha participado en el programa Ayuda a la Dislexia en el curso 2018-2019. De los 107 colegios que formaron parte del proyecto, 100 eran públicos, y, por tanto, restringimos la muestra a alumnos matriculados en colegios de esta titularidad.

Como se observa en el Cuadro 4, contamos con 28.719 observaciones para la evaluación en Lengua, 28.745 en Matemáticas y 28.716 en Inglés. Los resultados encontrados para las variables de control son los esperados por la literatura (Hanushek & Woessmann, 2011). A mayor nivel educativo del padre y de la madre, mejores son los resultados de los alumnos. La categoría omitida que sirve de referencia es «no fue al colegio». La dirección y magnitud de los coeficientes de correlación de estudios universitarios de Diplomatura, Grado, Licenciatura o Máster en los resultados de los alumnos de 3.º y 6.º de primaria son similares para ambos progenitores. Es interesante comprobar que, cuando se desagregan las estimaciones entre chicas y chicos, el efecto del nivel educativo alcanzado por el padre en los hijos es más elevado que en las hijas tanto para Matemáticas como para Inglés. Lo contrario ocurre con las madres, en donde el impacto de sus estudios es más elevado en Matemáticas e Inglés en

sus hijas que en sus hijos, un resultado consistente con las conclusiones de González de San Román y De la Rica (2016).

El número de libros que hay en los hogares de los alumnos es uno de los indicadores que más correlación muestran con los resultados de los alumnos (Hanushek & Woessmann, 2011). Se trata, de una variable de control importante que se emplea en las pruebas internacionales de PISA, TIMSS o PIRLS y que capta factores relacionados también con el nivel socioeconómico de los padres y su influencia en el rendimiento académico. El Cuadro 3 muestra que, incluso después de controlar por el nivel de estudios de la madre y del padre, el número de libros en casa se asocia a unos mejores resultados en las pruebas externas y estandarizadas. La categoría omitida es la de los «alumnos que tienen en su hogar 10 o menos libros». El impacto de la disposición de libros es lineal en todos los casos. Pero la categoría «de 11 a 50 libros», una categoría más que la referente, se correlaciona con un 20 % de la desviación estándar más de puntuación en Lengua, Matemáticas o Inglés.

En relación con el programa Ayuda a la Dislexia, los resultados del Cuadro 4 muestran que los alumnos de los colegios que participaron en el programa en 2018-2019 tienen un mejor resultado en Inglés, pero no en Lengua ni en Matemáticas. Al desagregar por género observamos que no solo ambos géneros mejoran en Inglés, sino que las alumnas también lo hacen en Lengua. Las alumnas de los colegios que han participado en el proyecto obtienen un 4,39 % de la desviación estándar más en Lengua con respecto a aquellas alumnas en centros educativos no incluidos en el programa. No hay ninguna correlación significativa para los chicos en Lengua. En Matemáticas, el coeficiente asociado a las alumnas de colegios tratados asciende al 3,22 % de la desviación estándar en comparación con aquellas alumnas de centros no tratados, sin embargo, el coeficiente no llega a ser significativo. Por último, en Inglés tanto las chicas como los chicos de colegios incluidos en el programa de Ayuda a la Dislexia tienen mejores resultados que sus pares de los demás centros públicos. La asociación con el rendimiento en Inglés en el caso de las chicas es de un 7,34 % de la desviación estándar, mientras que en el caso de los chicos es del 4,63 %. En general, la correlación entre la participación en el proyecto y los resultados en las pruebas externas y estandarizadas en 2018-2019 son más fuertes y positivas para el caso de las chicas que de los chicos.

La mayor magnitud y significatividad para la muestra de alumnas se encuentra en línea con el resultado habitual de la literatura de economía de la educación. Iniciativas educativas influyen en mayor medida en las alumnas (Felfe *et al*., 2015). Es interesante comprobar que la asociación entre el programa de Ayuda a la Dislexia y los resultados de los alumnos es más fuerte para las competencias de comunicación lingüística, ya sea en Lengua o en Inglés, siendo esta última la de mayor magnitud. Parece lógico que un proyecto que refuerza las capacidades lectoescritoras esté más relacionado con las competencias en comunicación lingüística que con las Matemáticas, al menos en el corto plazo.

CUADRO 4

ASOCIACIÓN ENTRE LA PARTICIPACIÓN EN EL PROGRAMA AYUDA A LA DISLEXIA Y LOS RESULTADOS EN LAS PRUEBAS EXTERNAS Y ESTANDARIZADAS EN LA COMUNIDAD DE MADRID. CENTROS PÚBLICOS DE LA COMUNIDAD DE MADRID

Variables	Muestra completa			Alumnas			Alumnos		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Lengua	Matemát.	Inglés	Lengua	Matemát.	Inglés	Lengua	Matemát.	Inglés
Tratamiento	1,4342 (1,610)	1,6278 (1,550)	5,8335*** (1,540)	4,3910** (2,219)	3,2204 (2,176)	7,3417*** (2,157)	–0,9935 (2,244)	–0,0205 (2,185)	4,6302** (2,153)
% de alumnos inmigrantes en el centro	–0,0721 (0,313)	0,5244* (0,302)	–0,6555** (0,301)	–0,3449 (0,429)	0,3962 (0,420)	–1,2006*** (0,416)	–0,1079 (0,440)	0,7471* (0,429)	–0,3328 (0,426)
Educac. del padre (<EGB o ESO)	–0,3692 (8,064)	3,0239 (7,833)	2,7399 (7,757)	4,6839 (10,681)	–2,9846 (10,516)	3,0274 (10,387)	–0,1685 (11,769)	5,1709 (11,612)	5,1963 (11,409)
Educac. del padre (EGB o ESO)	11,2831 (8,007)	12,1959 (7,777)	17,9578** (7,702)	18,1773* (10,587)	9,4759 (10,422)	18,2251* (10,295)	9,7545 (11,705)	11,4956 (11,550)	20,4351* (11,346)
Educac, del padre (Bach, BUP, COU, FPI)	19,3441** (8,023)	17,9190** (7,792)	29,6356*** (7,718)	28,0402*** (10,602)	14,2291 (10,436)	32,6933*** (10,310)	16,6719 (11,733)	17,9162 (11,579)	29,9636*** (11,375)
Educac. del padre (FPII)	21,9545*** (8,127)	21,2245*** (7,892)	34,0807*** (7,816)	25,9637** (10,751)	14,9724 (10,583)	32,0502*** (10,455)	23,7570** (11,872)	23,3472** (11,712)	39,1247*** (11,506)
Educac. del padre (Diplomatura Univ.)	33,5531*** (8,172)	33,8511*** (7,936)	46,2818*** (7,859)	38,8814*** (10,816)	30,6481*** (10,647)	49,0364*** (10,518)	33,0986*** (11,932)	33,7644*** (11,771)	45,8665*** (11,562)
Educac. del padre (Licenciatura o Grado)	36,6567*** (8,127)	36,2002*** (7,892)	51,2490*** (7,816)	39,9166*** (10,753)	29,6388*** (10,585)	51,3320*** (10,457)	38,8397*** (11,870)	39,0878*** (11,710)	53,9389*** (11,503)
Educac. del padre (Máster Universitario)	33,1038*** (8,445)	36,2106*** (8,196)	54,2991*** (8,117)	39,9300*** (11,207)	29,1040*** (11,029)	52,3346*** (10,896)	32,3052*** (12,297)	39,4991*** (12,121)	59,2811*** (11,906)
Educac. del padre (Doctorado)	35,8367*** (9,009)	40,6273*** (8,738)	61,5574*** (8,650)	41,6813*** (12,013)	31,9949*** (11,817)	55,2947*** (11,664)	36,1487*** (13,052)	45,5623*** (12,853)	70,4948*** (12,625)
Educac. de la madre (<EGB o ESO)	13,0416 (8,173)	3,2640 (7,939)	–1,3441 (7,863)	6,4583 (11,844)	13,9703 (11,672)	5,1463 (11,530)	13,5844 (10,934)	–2,2851 (10,769)	–9,4761 (10,582)
Educac. de la madre (EGB o ESO)	21,5925*** (8,033)	11,1801 (7,802)	3,6863 (7,725)	15,3055 (11,676)	21,5119* (11,504)	7,8663 (11,360)	22,9165** (10,719)	5,3568 (10,556)	–1,3619 (10,369)

CUADRO 4 *(Continuación)*

ASOCIACIÓN ENTRE LA PARTICIPACIÓN EN EL PROGRAMA AYUDA A LA DISLEXIA Y LOS RESULTADOS EN LAS PRUEBAS EXTERNAS Y ESTANDARIZADAS EN LA COMUNIDAD DE MADRID. CENTROS PÚBLICOS DE LA COMUNIDAD DE MADRID

	Muestra completa			Alumnas			Alumnos		
Variables	**(1) Lengua**	**(2) Matemát.**	**(3) Inglés**	**(4) Lengua**	**(5) Matemát.**	**(6) Inglés**	**(7) Lengua**	**(8) Matemát.**	**(9) Inglés**
Educac. de la madre (Bach, BUP, COU, FPI)	28,2359*** (8,035)	17,4384** (7,804)	14,7992* (7,730)	24,0583** (11,672)	28,7016** (11,500)	18,8007* (11,361)	26,6695** (10,728)	10,9800 (10,566)	9,4275 (10,380)
Educac. de la madre (FPII)	26,8540*** (8,140)	21,3501*** (7,906)	14,0209* (7,829)	21,5772* (11,818)	32,0096*** (11,644)	20,7676* (11,501)	28,3301*** (10,874)	14,4018 (10,706)	7,6374 (10,516)
Educac. de la madre (Diplomatura Univ.)	41,1554*** (8,143)	35,5687*** (7,907)	27,4245*** (7,831)	34,9418*** (11,814)	47,0286*** (11,638)	35,1225*** (11,496)	42,7759*** (10,883)	28,3798*** (10,716)	19,3878* (10,527)
Educac. de la madre (Licenciatura o Grado)	45,3107*** (8,115)	37,6426*** (7,881)	36,7567*** (7,805)	41,4755*** (11,778)	50,7779*** (11,603)	44,8952*** (11,462)	45,6806*** (10,844)	28,1986*** (10,679)	28,9801*** (10,490)
Educac. de la madre (Máster Universitario)	46,1614*** (8,399)	37,6956*** (8,153)	37,0518*** (8,074)	34,8259*** (12,154)	46,9712*** (11,971)	42,1049*** (11,825)	53,0943*** (11,253)	32,3465*** (11,074)	31,8349*** (10,877)
Educac. de la madre (Doctorado)	51,2058*** (9,018)	44,4888*** (8,747)	43,4555*** (8,662)	42,9424*** (12,927)	60,3737*** (12,727)	52,1611*** (12,576)	54,4500*** (12,190)	32,9030*** (11,985)	34,3759*** (11,767)
De 11 a 50 libros	19,5587*** (2,563)	18,4174*** (2,478)	20,9258*** (2,453)	22,1948*** (3,599)	17,4280*** (3,541)	25,4897*** (3,490)	15,6776*** (3,518)	20,1686*** (3,439)	16,1388*** (3,382)
De 51 a 100 libros	32,4271*** (2,612)	27,0795*** (2,524)	30,3026*** (2,499)	32,8463*** (3,649)	23,3145*** (3,589)	33,7433*** (3,538)	28,8804*** (3,602)	32,1228*** (3,520)	25,5297*** (3,463)
De 101 a 200 libros	40,9268*** (2,755)	33,3923*** (2,662)	38,0640*** (2,636)	43,0529*** (3,838)	29,5446*** (3,775)	42,2080*** (3,722)	35,5920*** (3,808)	38,5510*** (3,721)	32,2991*** (3,660)
Más de 200 libros	48,1152*** (2,776)	43,8810*** (2,683)	42,6753*** (2,657)	52,5959*** (3,865)	39,4923*** (3,802)	47,1861*** (3,750)	40,5821*** (3,841)	49,7608*** (3,753)	36,7545*** (3,690)
Constante	413,6064*** (8,463)	417,2208*** (8,251)	415,1718*** (8,172)	429,1933*** (11,602)	403,8580*** (11,495)	418,2019*** (11,356)	400,7198*** (12,004)	428,3189*** (11,857)	411,6298*** (11,648)
Observaciones	28.719	28.745	28.716	13.985	13.985	13.987	14.731	14.757	14.726
R-cuadrado	0,081	0,083	0,112	0,088	0,081	0,126	0,083	0,089	0,106

NOTAS: Muestra: alumnos de 3.º y 6.º de primaria evaluados en las pruebas externas de la Comunidad de Madrid en el curso 2018-2019. Errores estándar en paréntesis: *** p<0,01. ** p<0,05. * p<0,1.

FUENTE: Elaboración propia.

El inglés como segunda lengua es una asignatura clave para los alumnos que asisten a colegios bilingües, en los cuales asignaturas como Ciencias, Sociales, Educación Física, Música o Artes se imparten en dicho idioma. Una de las críticas que habitualmente reciben los programas bilingües existentes en varias comunidades autónomas, entre ellas Madrid, es la posible desventaja en rendimiento que la enseñanza en inglés puede suponer para el alumnado con necesidades educativas de apoyo específico, como es el caso de los estudiantes con dislexia. Los primeros resultados, que muestran correlación y no causalidad, estarían apuntando a que el programa Ayuda a la Dislexia se asocia a un mejor resultado en inglés de los alumnos, y con ello, podrían estar beneficiando en el rendimiento de la segunda lengua y en las asignaturas que se imparten en ese idioma. Esta mayor correlación para el inglés puede deberse a un factor de transferencia de conocimiento entre lengua nativa y segunda lengua previamente estudiado en dislexia. Contar con madurez en la adquisición de la lengua materna, concretamente tener la conciencia fonológica y la conciencia morfología más afianzadas –conciencias que específicamente trabaja Dytective U– tiene un impacto positivo en la adquisición del inglés (Siegel, 2008; Hernández García *et al*., 2018).

5. Conclusiones

Evaluamos el proyecto de Ayuda a la Dislexia de la Comunidad de Madrid dirigido a alumnos con dificultades lectoescritoras en el curso 2018-2019. El programa consiste en dos herramientas: *a*) una prueba a través de un *software* para comprobar los alumnos que presentan dificultades de lectoescritura y, por tanto, posible riesgo de dislexia y *b*) hasta 42.000 ejercicios en una plataforma digital para trabajar y reforzar las capacidades lectoescritoras de los estudiantes con peores resultados en el cribado. Un 7,55 % (n = 1,022) de los alumnos que finalizaron la prueba del programa correctamente (n = 13.406) presentaban dificultades lectoescritoras y, por tanto, riesgo de dislexia. Los resultados del análisis de correlación entre la participación del centro en el programa de Ayuda a la Dislexia y el rendimiento de los alumnos de esos colegios en las pruebas externas y estandarizadas de la LOMCE en 3.º y 6.º de primaria, muestran una asociación positiva y significativa con el rendimiento de Inglés.

Estos resultados apuntan a que el programa de Ayuda a la Dislexia no solo se asocia a una mejora en el idioma nativo sino también a la segunda lengua siendo especialmente positivo para las alumnas, cuyo diagnóstico se hace más complejo por diferencias de género en relación con la inhibición y rendimiento académico (Arnett *et al*., 2017) pero que gracias al *software* de la prueba de cribado son detectadas al mismo nivel que sus homólogos masculinos (Rello *et al*., 2020). Una de las ventajas del *software* es que diagnostica a niñas y a niños por igual. Las niñas son más difíciles de diagnosticar por los sistemas educativos, porque su rendimiento promedio en lectura es más elevado y pueden pasar más desapercibidas. Una hipótesis en la que trabajaremos como línea de investigación futura es si las chicas se benefician más

de Dytective y Dytective U porque son diagnosticadas y tratadas en mayor medida que los chicos. El análisis descriptivo de este artículo muestra una primera evidencia correlacional de que el programa de Ayuda a la Dislexia diagnostica más a las niñas que a los niños.

No obstante, se ha comprobado que los centros que participan en esta intervención tienen mejores resultados que los centros no tratados y que sus alumnos provienen de un entorno socioeconómico más elevado. Estas diferentes características de los colegios tratados podrían explicar su participación, que era voluntaria, en el programa de Ayuda a la Dislexia.

La investigación futura de este proyecto de evaluación de un programa público plantea realizar un análisis en la línea del *paper* de Machin *et al*. (2018) para explotar a través de un análisis de diferencias-en-diferencias que ha habido una nueva convocatoria de Ayuda a la Dislexia después del curso 2018-2019. De este modo los colegios que formaron parte de esta iniciativa en 2018-2019 se compararán con los que han entrado en el programa en el curso 2021-2022.

Referencias bibliográficas

Angrist, J. D., & Pischke, J. S. (2014). *Mastering Metrics: The Path from Cause to Effect*. Princeton University Press. https://ideas.repec.org/b/pup/pbooks/10363.html

Arnett, A. B., Pennington, B. F., Peterson, R. L., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2017). Explaining the sex difference in dyslexia. *Journal of Child Psychology and Psychiatry*, *58*(6), 719-727.

Beg, S., Lucas, A. M., Halim, W., & Saif, U. (2021). Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not. *American Economic Journal: Economic Policy*. https://www.aeaweb.org/articles?id=10.1257/pol.20200713

Carrillo-Gallego, M. S., Alegría, J., Miranda, P., & Sánchez, N. (2011). Evaluación de la dislexia en la escuela primaria: Prevalencia en español. *Escritos de Psicología*, *4*(2), 35-44.

Felfe, C., Nollenberger, N., & Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of Population Economics*, *28*(2), 393-422.

Fryer, R. G., & Howard-Noveck, M. (2020). High dosage tutoring and reading achievement: Evidence from New York City. *Journal of Labor Economics*, *38*(2), 421-452.

Gao, Q., Wang, H., Mo, D., Shi, Y., Kenny, K., & Rozelle, S. (2018). Can reading programs improve reading skills and academic performance in rural China? *China Economic Review*, *52*, 111-125.

González de San Román, A., & De la Rica, S. (2016). Gender Gaps in PISA Test Scores. The Impact of Social Norms and the Mother's Transmission of Role Attitudes. *Estudios de Economía Aplicada*, *34*(1), 79-108.

Goux, D., Gurgand, M., & Maurin, E. (2017). Reading enjoyment and reading skills: Lessons from an experiment with first grade children. *Labour Economics*, *45*, 17-25.

Guryan J, Kim J. S., & Park, K. H. (2016). Motivation and incentives in education: Evidence from a summer reading experiment. *Economics of Education Review*, *55*, 1-20.

Hanushek, E. A., & Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education, vol. 3*, (pp. 89-200). North Holland.

Hernández-García, F., Hernández-Pallarés, L. A., Valencia-García, M. T., Ramírez-Lajarín, F. J., & Abril-López, M. A. (2018). *Guía de dislexia del inglés para alumnos con dislexia y otras dificultades*. Consejería de Educación, Juventud y Deportes de la Región de Murcia.

Jiménez, J. E., de la Cadena, C. G., Siegel, L. S., O'Shanahan, I., García, E., & Rodríguez, C. (2011). Gender ratio and cognitive profiles in dyslexia: A cross-national study. *Reading and writing*, *24*, 729-747.

Jiménez, J. E., Guzmán, R., Rodríguez, C., & Artiles, C. (2009). Prevalencia de las dificultades específicas de aprendizaje: La dislexia en español. *Anales de Psicología/Annals of Psychology*, *25*(1), 78-85.

Lavy, V., Kott, A., & Rachkovski, G. (2021). Does Remedial Education at Late Childhood Pay Off After All? Long-Run Consequences for University Schooling, Labor Market Outcomes and Inter-Generational Mobility. *Journal of Labor Economics*.

Machin, S., McNally, S., & Viarengo, M. (2018). Changing How Literacy Is Taught: Evidence on Synthetic Phonics. *American Economic Journal: Economic Policy*, *10*(2), 217-241.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review*, *109*(4), 1426-1460.

Rello, L. (2018). *Superar la dislexia*. Paidós.

Rello L., Baeza-Yates R., Ali A., Bigham, J. P., & Serra, M. (2020). Predicting risk of dyslexia with an online gamified test. *PLOS ONE*, *15*(12). https://doi.org/10.1371/journal.pone.0241687

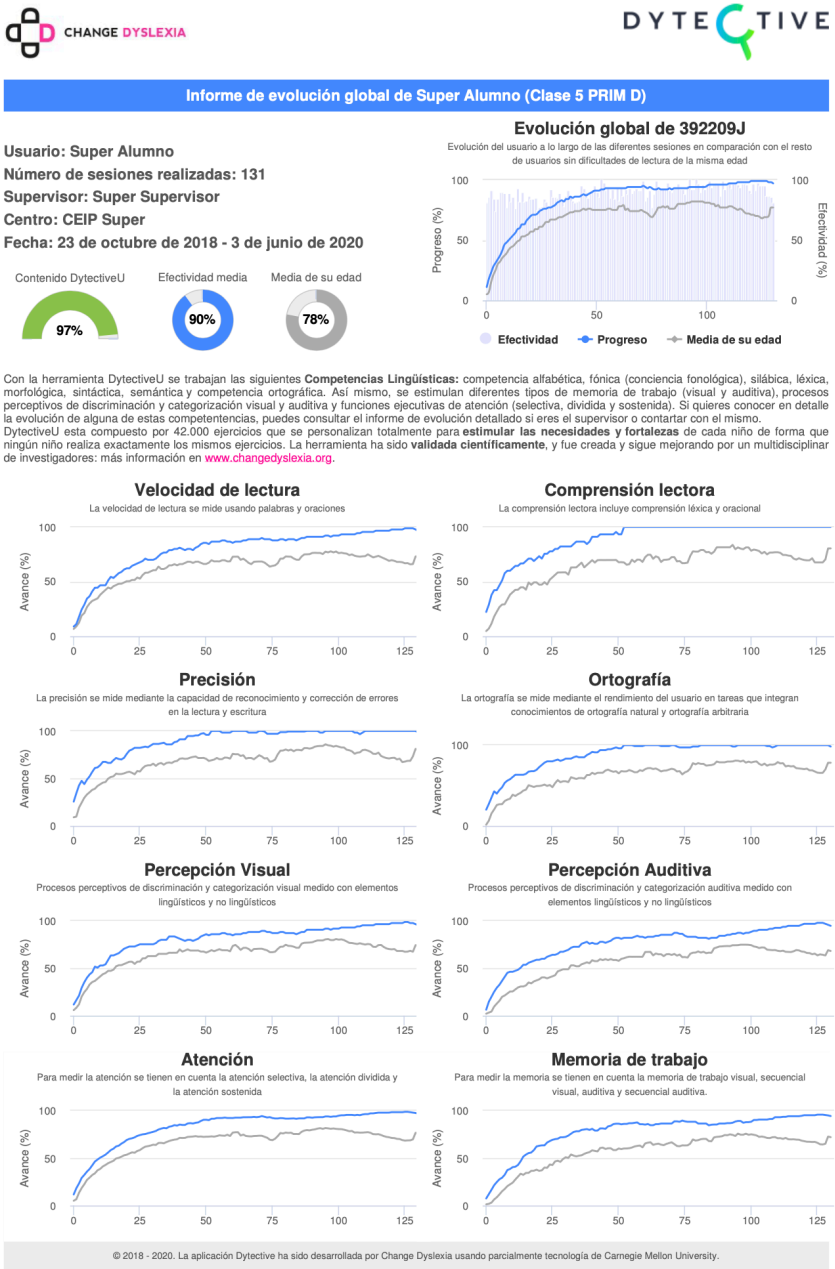
Shaywitz, S. E., & Shaywitz, B. A. (2003). Dyslexia (specific reading disability). *Pediatrics in Review*, *24*(5), 147-153.

Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders*, *28*(1), 15-27.

APÉNDICE

FIGURA A1

EJEMPLO DE INFORME DE SEGUIMIENTO DEL ALUMNO DE LA HERRAMIENTA DYTECTIVE U



FUENTE: Elaboración propia.

MINISTERIO DE INDUSTRIA, COMERCIO Y TURISMO
ICE
INFORMACIÓN COMERCIAL ESPAÑOLA
CUADERNOS ECONÓMICOS

NORMAS DE PUBLICACIÓN

Los manuscritos que se envíen para publicar en Cuadernos Económicos de ICE deberán ser originales, no publicados, ni en proceso de evaluación por otra publicación. Se aceptan trabajos en castellano o en inglés. Todos los trabajos recibidos serán sometidos para su aceptación a evaluación externa anónima.

1. Los originales se enviarán en formato Microsoft Word a la dirección de correo electrónico, revistasice.sscc@mincotur.es o directamente a la editora, María del Carmen Díaz Roldán (carmen.diazroldan@uclm.es), señalando en el asunto «Cuadernos Económicos de ICE». También pueden enviarse a través de la página web de *Revistas ICE*.
2. La extensión del trabajo, incluyendo cuadros, gráficos, tablas, notas y referencias bibliográficas, no excederá de 30 páginas. La fuente será Times New Roman 12, márgenes de 3 cm, espaciado doble y paginado en la parte inferior derecha.
3. Cada original incluirá en una primera página independiente el título del artículo y la fecha de conclusión del mismo; el nombre y apellidos de los autores, afiliación institucional, dirección, teléfono y correo electrónico de cada uno de ellos.
4. En la primera página del texto se incluirá:
   - El título en castellano y en inglés.
   - Un resumen del trabajo, en castellano y en inglés, con una extensión que no exceda de 100 palabras.
   - Una lista de 2 a 6 palabras clave
   - Los códigos (entre 1 y 5) de la clasificación del Journal of Economic Literature (JEL) que puede consultarse en el siguiente enlace https://www.aeaweb.org/jel/guide/jel.php
5. En las siguientes páginas se incorporará el texto, la información gráfica y la bibliografía con la siguiente organización:
   - Se incluirá, por este orden, introducción, desarrollo, conclusiones y referencias bibliográficas. Si hubiera anexos, se insertarán tras las referencias bibliográficas y deberán llevar título.
   - En un archivo Excel independiente se incluirá la representación gráfica (cuadros, gráficos, diagramas, figuras, etcétera), que debe llevar título, estar numerada y referenciada en el texto. En la parte inferior se incluirán la fuente de información y, en su caso, las notas aclaratorias.
   - Los apartados y subapartados se numerarán en arábigos según el siguiente modelo:

1. Título del apartado (primer nivel)

1.1. Subapartado (segundo nivel)

1.1.1. Sección del subapartado (tercer nivel)

- Las notas a pie de página irán integradas en el texto y su contenido debe estar al final de su misma página en tamaño 10, Times New Roman.
- En las citas en el texto, de acuerdo con las normas de la American Psychological Association (Normas APA), se indicará entre paréntesis el apellido del autor y el año. Ej.: (Martínez, 1991).
- Las referencias a siglas deben ir acompañadas, en la primera ocasión en que se citen, de su significado completo.
- Al final del texto se recogerá la bibliografía utilizada, bajo el título de «Referencias bibliográficas», ordenada alfabéticamente según las Normas APA. A continuación se recogen los ejemplos de referencias más utilizados:

Libro

Apellido, A. A. (Año). *Título*. Editorial.
Pilling, D. (2019). *El delirio del crecimiento*. Taurus.

Libro electrónico con DOI (o URL)

Apellido, A. A. (Año). *Título*. https://doi.org/xxx
Nicolleti, B. (2017). The Future of Fintech: Integrating Finance and Technology in Financial Services. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-51415-4

Capítulo de libro

Apellido, A. A. (Año). Título del capítulo o la entrada. En C. C. Apellido (Ed.), *Título del libro* (pp. xx-xx). Editorial.
Demas, M., & Agnew, N. (2013). Conservation and sustainable development of archaeological sites. En I. Rizzo y A. Mignosa, *Handbook on the Economics of Cultural Heritage* (pp. 326-343). Edward Elgar.

Publicaciones periódicas formato impreso

Apellido, A. A. (Año). Título del artículo. *Nombre de la revista*, *volumen* (número), pp-pp.
Fernández-Blanco, V., Prieto-Rodríguez, J., & Suárez-Pandiello, J. (2017). A quantitative analysis of reading habits in Spain. *International Journal of Arts Management, 19*(3), 19-32.

Publicaciones periódicas online con DOI (o URL)

Apellido, A. A. (Año). Título del artículo. *Nombre de la revista*, *volumen*(número), pp-pp. https://doi.org/xxx
Amador, J., & Cabral, S. (2017). Networks of Value-added Trade. *The World Economy, 40*(7), 1265-1512. https://doi.org/10.1111/twec.12469

Informe oficial en web

Organismo. (Año). *Título del informe*. http://www…

Fondo Monetario Internacional. (2019). *Global Financial Stability Report*. https://www.imf.org/en/Publications/GFSR/Issues/2019/10/01/global-financial-stability-report-october-2019

Ley/Reglamento

Título de la ley. *Publicación*, número, fecha de publicación, pp-pp. http://www.boe.es....

Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa. *Boletín Oficial del Estado*, nº. 295, de 10 de diciembre de 2013, pp. 97858 a 97921. https://www.boe.es/buscar/pdf/2013/BOE-A-2013-12886-consolidado.pdf

Orden de la lista de referencias bibliográficas

Las referencias se ordenan alfabéticamente y, en caso de varios trabajos realizados por el mismo autor/a, el criterio es el siguiente:

- Primero los trabajos en los que el/la autor/a figura solo/a. Correlativos de año más antiguo al año más actual de publicación.
- Segundo, aquellos trabajos colectivos en los que el/la autor/a es el/la primero/a. Correlativos por fecha.
- Tercero, en caso de coincidencia exacta de autores y fechas, debe citarse cada trabajo añadiendo una letra a la fecha. Ej.: 2014a, 2014b, etc.

En la página web del grupo de revistas de Información Comercial Española se publican los artículos en formato electrónico y se recogen los principios editoriales de la publicación: ***www.revistasICE.com***

CUADERNOS ECONÓMICOS DE INFORMACIÓN COMERCIAL ESPAÑOLA (CICE) ISSN 0210-2633

SUSCRIPCIÓN ANUAL

CUADERNOS ECONÓMICOS DE INFORMACIÓN COMERCIAL ESPAÑOLA (2 NÚMEROS/AÑO)			
	ESPAÑA 1 año	UNIÓN EUROPEA 1 año	RESTO DEL MUNDO 1 año
SUSCRIPCIÓN	30,00 €	30,00 €	30,00 €
Gastos de envío España	1,12 €	1,60 €	2,08 €
Más 4% de IVA Excepto Canarias, Ceuta y Melilla	1,24 €		
TOTAL	32,36 €	31,60 €	42,08 €

EJEMPLARES SUELTOS

CUADERNOS ECONÓMICOS DE INFORMACIÓN COMERCIAL ESPAÑOLA			
	ESPAÑA 1 ejemplar	UNIÓN EUROPEA 1 ejemplar	RESTO DEL MUNDO 1 ejemplar
NÚMERO SUELTO	18,00 €	18,00 €	18,00 €
Gastos de envío España	0,56 €	0,80 €	1,04 €
Más 4% de IVA Excepto Canarias, Ceuta y Melilla	0,74 €		
TOTAL	19,30 €	18,80 €	19,04 €

Suscripciones y venta de ejemplares sueltos
Ministerio de Industria, Comercio y Turismo. Centro de Publicaciones.
C/ Panamá, 1. Planta 0, despacho 3.
Teléfonos: 91 3 49 51 29 – 91 3 49 51 33 (venta de publicaciones)
91 3 49 43 35 (suscripciones)
Correo electrónico: CentroPublicaciones@mincotur.es

GOBIERNO DE ESPAÑA
MINISTERIO DE INDUSTRIA, COMERCIO Y TURISMO
SECRETARÍA GENERAL TÉCNICA
SUBDIRECCIÓN GENERAL DE DESARROLLO NORMATIVO, INFORMES Y PUBLICACIONES
CENTRO DE PUBLICACIONES

MINISTERIO DE INDUSTRIA,
COMERCIO Y TURISMO

CUADERNOS ECONÓMICOS
ICE
INFORMACIÓN COMERCIAL ESPAÑOLA

CUADERNOS ECONÓMICOS

ÚLTIMOS NÚMEROS PUBLICADOS

63. Aspectos del mercado de trabajo español
64. Crisis y reformas de los sistemas de Seguridad Social: Temas a debate (I)
65. Crisis y reformas de los sistemas de Seguridad Social: Temas a debate (II)
66. Microeconometría y conducta: Aplicaciones para el caso español
67. La economía de la salud
68. Modelos de microsimulación
69. Instrumentos derivados
70. Nuevas tendencias en Historia Económica
71. Diseño y evaluación de políticas de protección ambiental
72. Decisiones de inversión y adopción de nuevas tecnologías
73. La gestión de la actividad innovadora en las empresas
74. Capital humano y empleo
75. Desigualdad, equidad y eficiencia en salud y servicios sanitarios
76. Contribuciones recientes a la Economía del Desarrollo
77. Contribuciones a la Economía Experimental
78. Aportaciones recientes a la Economía Industrial
79. Contribuciones a la economía de la energía
80. Evaluación económica de infraestructuras de transporte
81. Análisis económico de las telecomunicaciones en España
82. Comercio internacional, empresas y competitividad
83. Economía de las energías renovables
84. Mejorar la productividad: un reto pendiente en la economía española
85. Temas actuales de Economía Política
86. Gestión ambiental de las organizaciones
87. Inmigración y crisis económica en España
88. Defensa de la Competencia en España: un análisis sectorial
89. Contribuciones a la Economía de la Innovación: análisis de la I+D+i empresarial
90. La vivienda tras la crisis financiera global: nueva visión, viejos problemas
91. Política económica europea
92. Una panorámica del sector bancario después de la crisis
93. Turismo sostenible: un debate todavía abierto a la discusión
94. Sector exterior y economía internacional
95. Capital humano y mercado laboral
96. Sistema Nacional de Salud: retos y tareas pendientes tras la Gran Recesión
97. ¿Economía colaborativa?: antecedentes, situación y desafíos en su regulación
98. Economía de la Cultura
99. Economía del comportamiento
100. Globalización y sostenibilidad
101. Indicadores macroeconómicos ambientales. El caso español
102. La evaluación de las políticas públicas en España: recursos, metodologías y algunos ejemplos

CUADERNOS ECONÓMICOS

ICE

INFORMACIÓN COMERCIAL ESPAÑOLA

GOBIERNO DE ESPAÑA

MINISTERIO DE INDUSTRIA, COMERCIO Y TURISMO

9 770210 263007 00102

P.V.P.: 18 € + IVA