

Gestión y Análisis de Políticas Públicas, número 35, julio de 2024
Sección: ARTÍCULOS
Recibido: 20-09-2023
Modificado: 13-11-2023
Aceptado: 21-11-2023
Publicación anticipada: 12-12-2023
Publicado: 01-07-2024
ISSN: 1989-8991 – DOI: <https://doi.org/10.24965/gapp.11259>
Páginas: 7-25



Referencia: Junquera, Á. F. (2024). Efectos de la asistencia en la búsqueda de empleo: una revisión sistemática para España. *Gestión y Análisis de Políticas Públicas*, 35, 7-25. <https://doi.org/10.24965/gapp.11259>

Efectos de la asistencia en la búsqueda de empleo: una revisión sistemática para España

Effects of job search assistance: A systematic review for Spain

Junquera, Álvaro F.

Universitat Autònoma de Barcelona. Institut d'Estudis del Treball, Departamento de Sociología (España – Spain)

ORCID: <https://orcid.org/0000-0003-0080-7608>

alvaro.junquera@uab.cat

NOTA BIOGRÁFICA

Investigador predoctoral en el Institut d'Estudis del Treball de la Universitat Autònoma de Barcelona, con una ayuda FPU del Ministerio de Universidades. Es doctorando en Sociología por la Universitat Autònoma de Barcelona y en Economía por la Università degli Studi di Padova. Sus intereses de investigación son la evaluación de políticas de empleo y el desarrollo de modelos de perfilado.

RESUMEN

Objetivos: este artículo realiza una revisión de la literatura sobre programas de asistencia en la búsqueda de empleo en España. Nos preguntamos qué intervenciones de este tipo han sido evaluadas, qué métodos se han usado y cuál es el efecto promedio que generan sobre el empleo y el empleo de larga duración. **Metodología:** para ello, hemos construido un conjunto de datos de 1042 estimaciones de impacto a partir de veintidós estudios. El análisis de los datos combinó el análisis cuantitativo bivariado con una síntesis narrativa a partir de la mejor evidencia disponible. **Resultados:** respecto al objeto y al método de estudio, los resultados indican una notable concentración de la evaluación en intervenciones profundamente heterogéneas y en estrategias de identificación con asunciones demasiado fuertes como para ser creíbles. Respecto al impacto de la intervención, centrándonos en las estimaciones con mayor validez interna, el efecto promedio sobre el empleo es positivo, pero de baja magnitud (2,71%). **Conclusiones:** nuestro trabajo señala lagunas de la literatura y estrategias de identificación más creíbles incluso con los datos administrativos normalmente disponibles. Además, indicamos qué intervenciones cuentan con evidencia sólida como para ser consideradas buenas prácticas.

PALABRAS CLAVE

Políticas activas de empleo; asistencia en la búsqueda de empleo; servicios públicos; revisión de literatura; evaluación de políticas públicas.

ABSTRACT

Objectives: This article reviews the literature on job search assistance programmes in Spain. We ask what interventions of this type have been evaluated, what methods have been used and what is the average effect they generate on employment and long-term employment. **Methodology:** For this purpose, we constructed a dataset of 1042 impact estimates from 22 studies. The data analysis combined quantitative bivariate analysis with a narrative synthesis from the best available evidence. **Results:** Regarding the object and method of study, the results indicate a remarkable concentration of the evaluation on deeply heterogeneous interventions and on identification strategies with assumptions too strong to be credible. Regarding the impact of the intervention, focusing on the most internally valid estimates, the average effect

on employment is positive, but of low magnitude (2,71 %). **Conclusions:** Our work shows the possibility of relying on techniques that facilitate the transportability of results and on more credible identification strategies even with the administrative data normally available. Furthermore, we point out which interventions have solid evidence to be considered good practices to be implemented in nearby contexts.

KEYWORDS

Active labour market policies; job search assistance; public services; literature review; public policy evaluation.

SUMARIO

1. INTRODUCCIÓN. 2. LA ASISTENCIA EN LA BÚSQUEDA DE EMPLEO. 3. METODOLOGÍA. 3.1. PRODUCCIÓN DE DATOS. 3.2. ANÁLISIS DE DATOS. 4. RESULTADOS. 4.1. LOS PROGRAMAS EVALUADOS. 4.2. LOS EFECTOS DE LOS PROGRAMAS. 5. DISCUSIÓN. 5.1. SOBRE LA VARIABLE DE TRATAMIENTO. 5.2. SOBRE LA VARIABLE RESPUESTA. 5.3. SOBRE LA ESTRATEGIA ANALÍTICA. 6. CONCLUSIÓN. AGRADECIMIENTOS. FINANCIACIÓN. REFERENCIAS BIBLIOGRÁFICAS.

1. INTRODUCCIÓN

El desempleo es uno de los principales problemas sociales en España. Para combatirlo, los Gobiernos suelen aprobar políticas activas de empleo (en adelante, PAE). Sin embargo, las buenas intenciones no siempre se traducen en buenas consecuencias. Por ello, los diseñadores de intervenciones necesitan conclusiones científicas sobre cuáles son las herramientas más efectivas para abordar el paro.

Este trabajo tiene como objetivo revisar la evidencia empírica acerca de los efectos sobre el empleo en España de los programas de asistencia en la búsqueda de empleo (en adelante, ABE), un subtipo de PAE. La justificación de nuestra investigación es doble. Por un lado, la ABE ha recibido menos atención por parte de los investigadores del mercado laboral (Card *et al.*, 2018). Sin embargo, en 2021 se gastaron 731 millones de euros de fondos públicos y más de 372000 personas participaron en acciones de ABE con financiación pública en toda España (DG EMPL, 2023). Por otro lado, restringimos nuestro alcance a España para relajar algunos supuestos de estudios previos. Los metaanálisis internacionales suelen descartar una interacción entre las PAE y características de los países, como la legislación de protección del empleo. En este artículo, comparamos estimaciones que comparten gran parte del mismo marco institucional, ajustando así por una posible interacción entre el efecto de las PAE y características propias del contexto español.

Concretamente, abordamos las siguientes preguntas de investigación: (1) ¿qué programas de ABE implementados en España han sido evaluados?; (2) ¿cómo han sido evaluados estos programas?, y (3) ¿cuál es el efecto de la ABE sobre el empleo y sobre el empleo de larga duración? Para responder a estas preguntas, hemos construido una nueva base de datos a partir de veintidós estudios, consiguiendo un total de 1042 estimaciones de impacto. En primer lugar, hemos clasificado las estimaciones según factores de diseño, metodológicos y contextuales. En segundo lugar, hemos resumido la evidencia poniendo un mayor énfasis en los artículos metodológicamente más sólidos.

Los resultados más importantes fueron los siguientes. Primero, las intervenciones más evaluadas son programas altamente heterogéneos, bien por su adaptación al participante (itinerarios) o bien por la agregación de diversas acciones (donde es difícil identificar los mecanismos importantes). Segundo, el 80 % de los estudios que analizan el efecto sobre el empleo lo hacen a través de la estrategia de identificación con más dificultades para corregir el sesgo de selección: la selección bajo observadas. Tercero, considerando solo los análisis con alta validez interna, la estimación del efecto promedio de la ABE sobre el empleo es poco concluyente (positiva, pero del 2,71 %).

Estos resultados generan contribuciones tanto en la dimensión básica como aplicada de la investigación. En el ámbito de investigación de las PAE, este es el primer trabajo que revisa de manera comprehensiva la literatura sobre los efectos de la ABE en España. Los trabajos anteriores, o bien no han sido exhaustivos o bien no han incluido estudios escritos en una lengua diferente al inglés. En el ámbito aplicado, nuestro estudio indica a las diferentes organizaciones proveedoras de servicios de ABE en España cuáles son las intervenciones que parecen generar efectos positivos de acuerdo con evidencia sólida.

El artículo continúa con una síntesis de los antecedentes sobre los efectos de la ABE en el empleo (sección 2). En la sección 3 describimos los datos y las técnicas utilizadas. En la sección 4 presentamos los hallazgos de nuestra investigación. La sección 5 comenta algunos inconvenientes de nuestro conjunto de datos con idea de mejorar las futuras evaluaciones. Por último, la sección 6 aporta algunas conclusiones.

2. LA ASISTENCIA EN LA BÚSQUEDA DE EMPLEO

El desempleo puede abordarse con políticas dirigidas a los mercados de factores o a los mercados de productos. En el primer capítulo, las PAE tienen una posición reconocida. Se suelen englobar bajo esta etiqueta cuatro tipos de políticas: políticas de búsqueda de empleo, de formación de adultos, de subsidios al empleo y de subsidios al emprendimiento (Malo, 2018; Caliendo y Künn, 2011). Las políticas de búsqueda de empleo suelen denominarse también con otros nombres, como servicios de empleo, asistencia en la búsqueda de empleo, orientación, asesoramiento, monitoreo... Sin embargo, tratar todos estos conceptos como sinónimos dificulta la detección de los mecanismos que podrían estar operando bajo los efectos encontrados.

Por ello, partimos de un marco conceptual en el que la «asistencia en la búsqueda de empleo» y los «requisitos sobre la búsqueda de empleo» son dos tipos diferentes de políticas de búsqueda de empleo. Esta distinción sigue la clasificación de Bonoli (2013) y otros artículos de la literatura como Brown y Koettl (2015) o Crépon y Van den Berg (2016). Por un lado, los programas de asistencia en la búsqueda de empleo «explican a los desempleados cómo buscar trabajo de manera efectiva [...] y pueden recomendar puestos de trabajo a los que el participante debería postularse» (Marinescu, 2017). Por otro lado, los programas de requisitos sobre la búsqueda de empleo imponen deberes a los participantes sobre las tareas de búsqueda de empleo.

En nuestra investigación, damos un paso más en la clasificación de las PAE. Para caracterizar los diferentes programas de ABE, distinguimos los diversos componentes principales que puede presentar una intervención. De este modo, ampliamos la taxonomía de Liu *et al.* (2014) con algunos componentes relacionados con la teoría de la búsqueda de empleo desarrollada en economía laboral (Cahuc *et al.*, 2014). Concretamente, estos nueve componentes se pueden dividir en tres grupos: desarrollo de habilidades, mejora de la motivación y suministro de información difícil. El gráfico 1 presenta esta taxonomía. Los componentes relacionados con el primer y segundo grupo fueron propuestos en Liu *et al.* (2014). El tercer grupo está integrado por componentes destinados a corregir fallos de información incompleta en los mercados laborales.

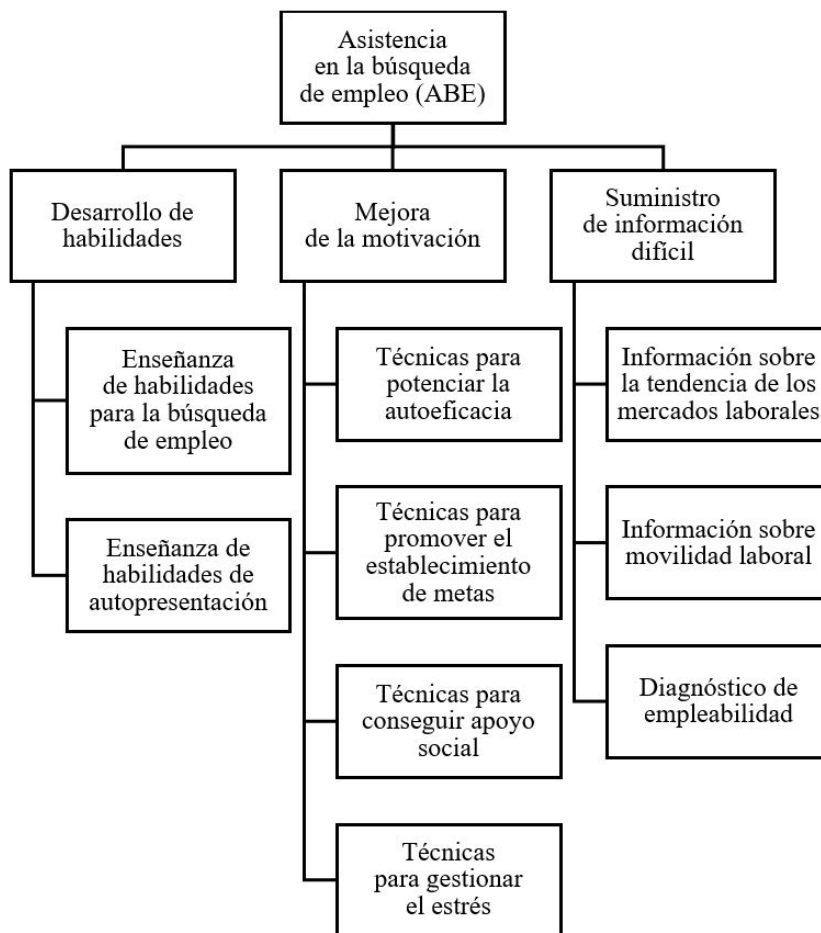
En primer lugar, tenemos dos componentes de desarrollo de habilidades¹. Por un lado, las habilidades para la búsqueda de empleo se refieren a aquellas competencias requeridas para identificar los canales adecuados para encontrar una vacante y las ocupaciones que encajan con el perfil del participante. Por otra parte, mejorar la autopresentación está relacionado con una adecuada candidatura a un puesto de trabajo tanto para momentos de comunicación síncrona (p. ej., la comunicación verbal y no verbal en una entrevista) como asíncrona (p. ej., la redacción de currículums o cartas de presentación).

En segundo lugar, encontramos los componentes de mejora de la motivación, que están íntimamente ligados a las teorías psicológicas sobre procesos cognitivos. Aquí incluimos diferentes actividades destinadas a fomentar la autoeficacia (como el aprendizaje vicario o la autorientación verbal de pensamientos), el establecimiento de objetivos (p. ej., la revisión del progreso o rutinas), el apoyo social (p. ej., con el apoyo de compañeros o familiares) o el manejo del estrés (con acciones como la redacción expresiva).

En tercer lugar, los componentes de información intentan ampliar el conjunto de ocupaciones consideradas o aumentar la intensidad de la búsqueda de empleo. Por un lado, la provisión de información sobre las tendencias del mercado laboral o sobre la movilidad laboral intenta ajustar la oferta y la demanda de mano de obra destacando la alta o baja demanda de trabajo en determinadas ocupaciones o zonas. Por otro lado, el diagnóstico de empleabilidad proporciona información al usuario sobre la probabilidad de encontrar empleo dadas sus circunstancias tanto en el presente como en el futuro. Esto es una reformulación de lo que Liu *et al.* (id.) llaman estímulo a la proactividad.

¹ Son políticas de búsqueda de empleo porque dotan al participante de las competencias necesarias para *conseguir* un empleo, mientras que las políticas de formación de adultos desarrollan habilidades para *ejecutar* las tareas asociadas a uno o más puestos de trabajo.

GRÁFICO 1. POSIBLES COMPONENTES DE LA ASISTENCIA EN LA BÚSQUEDA DE EMPLEO (ABE)



Fuente: elaboración propia.

En la literatura internacional se han analizado muchas de las evaluaciones de PAE en los trabajos de Vooren *et al.* (2019), Kluve *et al.* (2019) y Card *et al.* (2018). Sin embargo, estos estudios presentan algunos problemas en relación con la evidencia sobre los efectos de la ABE para el caso español. En primer lugar, la muestra de Card *et al.* (í.d.) únicamente incluye veintitrés estimaciones de PAE implementadas en los países del sur de la Unión Europea, mientras que la de Vooren *et al.* (2019) solo veinte. Concretamente, Card *et al.* (2018) solo incorporaron una estimación sobre políticas de búsqueda de empleo implementadas en España. Este problema está ligado al hecho de que solo incluyeron estudios redactados en inglés. En segundo lugar, en estas revisiones no hay una diferenciación clara entre intervenciones de asistencia y de requerimientos.

Vooren *et al.* (2019) encontraron que, en comparación con otros tipos de PAE, las políticas de búsqueda de empleo son las más efectivas a corto plazo. En la evaluación de Card *et al.* (2018), los autores concluyeron que estas políticas son en el corto plazo la segunda mejor opción para los parados de ocupaciones con bajos ingresos o con baja vinculación al mercado laboral. No obstante, cuando se ha considerado la posible existencia de externalidades negativas sobre los no tratados, los efectos netos son mucho menores (Cheung *et al.*, 2023). Más allá de las PAE, Liu *et al.* (2014) revisaron las evaluaciones de ABE publicadas hasta 2013 en inglés. Encontraron que aquellas acciones que mezclaron tanto desarrollo de habilidades como mejora de la motivación fueron más eficaces. Nuestra investigación pretende complementar la de Liu *et al.* (í.d.), aplicando gran parte de su marco conceptual al caso español.

En España, la evidencia empírica sobre los efectos de estos programas es más escasa. Debido al proceso de descentralización, las Administraciones autonómicas pueden promover la evaluación de muchos de estos programas. El Gobierno central también despliega algunas PAE, por lo que también se le ha reclamado

una mayor evaluación de sus acciones (Catalá-Pérez y del Pino, 2018). Aunque la competencia es compartida, las comunidades autónomas juegan un papel importante en el diseño de muchas de las intervenciones y en la gestión de los datos de participación en las PAE. En la última década, la financiación estatal de las PAE ha dependido, en mayor medida, de indicadores cuantitativos (Hernández-Moreno y Ramos Gallarín, 2017), pero estos están todavía lejos de recoger estimaciones de evaluaciones de impacto. No obstante, se han realizado algunas revisiones sobre las evaluaciones de PAE en España (Ramos *et al.*, 2010; Malo y Cueto, 2015; Arranz y García Serrano, 2022), pero estas presentan dos problemas sustanciales: no han sido exhaustivas (especialmente respecto a la ABE) y están desfasadas. Nuestro trabajo trata de mejorar ambos aspectos.

3. METODOLOGÍA

3.1. Producción de datos

Para responder a las preguntas de investigación, recopilamos un conjunto de evaluaciones de impacto de ABE realizadas en España. Esta tarea implicó buscar, transcribir y codificar datos. Previo a la búsqueda, delimitamos nuestro objeto de estudio de acuerdo al marco conceptual antes esbozado. Así, en cuanto a la muestra de estudios considerados, se determinaron los criterios de inclusión/exclusión de acuerdo con nuestro marco conceptual y con la naturaleza de nuestras preguntas de investigación.

En cuanto a la metodología del estudio, se estableció un solo criterio de inclusión/exclusión. Como nos preguntábamos por parámetros poblacionales con interpretación causal, solo incluimos análisis cuantitativos con técnicas experimentales o cuasiexperimentales. Por lo tanto, siguiendo los estándares actuales de la econometría (Cunningham, 2021), consideramos tanto los experimentos aleatorizados como los estudios observacionales con correcciones del sesgo de selección (como el emparejamiento, la regresión con variable instrumental, los efectos fijos, la regresión discontinua o el control sintético).

En cuanto al tratamiento de interés, también impusimos otro criterio de inclusión/exclusión: el estudio debería analizar una intervención en la que la ABE fuera la acción principal. Solo incluimos aquellas evaluaciones que estimaron el efecto sobre cualquiera de estas dos variables dependientes: (a) conseguir un empleo o (b) conseguir un empleo de larga duración. De forma flexible, consideramos como equivalente a la segunda opción tanto conseguir un contrato indefinido como uno de duración determinada, pero con una duración mínima de seis meses. Como se muestra en Cárdenas del Rey (2020), en España la duración media de los contratos temporales ha descendido fuertemente desde 2006. Debido a que la duración media en 2019 fue de algo más de cincuenta días, consideramos que seis meses es una duración relativamente larga.

Una vez fijados los criterios de inclusión/exclusión de los estudios, la primera etapa de recolección de datos fue una búsqueda a través de tres canales: buscadores académicos, búsqueda manual y contactos con expertos. En primer lugar, se realizaron diferentes búsquedas en la Web of Science y Google Scholar con las palabras clave más representativas para referirse a la ABE, tanto en español como en inglés. En segundo lugar, buscamos artículos concretos consultando la sección de referencias de los trabajos que ya habíamos recopilado. Durante este tiempo, también consideramos la *literatura gris* mediante la búsqueda de evaluaciones en las páginas web de diferentes diseñadores de PAE a nivel nacional y regional. Esta es una de las contribuciones de nuestro trabajo, ya que revisiones anteriores no incluyeron en su muestra la literatura gris. En tercer lugar, enviamos correos electrónicos a algunos profesores de disciplinas relacionadas con el estudio de las intervenciones en el mercado laboral y con la evaluación de programas.

La segunda etapa de la recopilación de datos comprendió la transcripción de los datos. Hemos diferenciado el número de estudios y el número de estimaciones de impacto. Cuando había diferentes estimaciones disponibles para el mismo horizonte de tiempo, incluimos hasta dos estimaciones de impacto, dando preferencia a aquellas en puntos intermedios o finales del intervalo de tiempo. Para permitir comparaciones entre evaluaciones, solo transcribimos estimaciones para submuestras creadas a partir de nuestras variables de interés (ver tabla 1)². En los casos en los que algunas estadísticas o partes del análisis no se

² Solo se incluye la estimación agregada de programas en el caso de que antes no se haya ofrecido una desagregación por alguna de las variables de interés del grupo diana (sexo, edad, desventaja).

describieron suficientemente, contactamos con los autores. Nótese que solo incluimos evaluaciones que estimaran el efecto de la ABE sobre las variables dependientes *obtener un empleo* u *obtener un empleo prolongado*.

La tercera etapa de la producción de datos, la codificación, consistió en dar valores a ciertas variables no reportadas directamente en los estudios considerados. Por un lado, trabajamos con variables contextuales, como la tasa de desempleo, pero también con dimensiones menos explícitas, como el componente principal de la intervención. La tabla 1 lista el conjunto de variables consideradas. En el apéndice puede encontrarse una definición detallada de cada una de ellas.

TABLA 1. VARIABLES CONSIDERADAS DURANTE LA REVISIÓN DE LA LITERATURA

Grupo de variables	Variable
Diseño del programa	Componente principal de la intervención
	Sexo del grupo diana analizado
	Edad del grupo diana analizado
	Desventaja del grupo diana analizado
Metodológico	Estrategia de identificación
	Variables ajustadas
	Tamaño de la muestra
	Horizonte temporal de la variable de respuesta
	Variable respuesta de dimensión stock o dimensión flujo
	Estimando
	Incluye un procedimiento de inferencia
	Publicado en una revista científica
Contexto de implementación	Territorio de implementación
	Año-trimestre de implementación
	Tasa de desempleo

Fuente: elaboración propia.

3.2. Análisis de datos

El análisis de datos se llevó a cabo considerando los dos objetivos específicos de nuestro trabajo. El primer objetivo específico era realizar una descripción exhaustiva de las evaluaciones de impacto de programas de ABE implementados en España. Para resumir la información de nuestra base de datos, realizamos un análisis cuantitativo bivariado a partir de tablas de frecuencia para cada variable dependiente. El segundo objetivo específico consistió en la estimación del efecto promedio de estas intervenciones. Para ello, empleamos la técnica de síntesis de la mejor evidencia (Slavin, 1995; Petticrew y Roberts, 2008). Este método considera las deficiencias que presentan otras técnicas como el metaanálisis cuando las muestras están formadas por muchas estimaciones sesgadas. Por ello, analiza de forma detallada solo aquellos artículos con los más altos estándares metodológicos. No obstante, fijar umbrales tan elevados durante la etapa de recolección de datos proporcionaría una imagen incompleta de la literatura.

Por tanto, tras conformar la muestra de estimaciones con los estándares habituales de la literatura econométrica, hemos realizado un segundo filtro considerando también las particularidades de nuestro objeto de estudio. Inspirados en Slavin (1995) y en Bell *et al.* (1995), el criterio para designar un estudio como perteneciente al grupo de mejor evidencia disponible fue la validez interna. Para definir la validez interna de una estimación, hemos aplicado la ley de la credibilidad decreciente de Manski (2007), la cual sostiene

que a mayor fuerza de las asunciones mantenidas, menor credibilidad de la inferencia. De este modo, para juzgar la validez interna de la estimación, hemos valorado las asunciones mantenidas sobre el sesgo de no-confusión.

Tal como muestra la tabla 2, esto lo hemos hecho en función del conocimiento y del uso que el investigador hace de la ecuación estructural (causal) de participación en el tratamiento. Las estrategias experimentales y de regresión discontinua tienen la ventaja de que, por el diseño de la intervención, conocen la ecuación estructural de participación en el tratamiento (donde participar viene determinado o por un sorteo o por una variable puntuación conocida). Dicho conocimiento permite asegurarnos de que, bien implementado y con un tamaño de muestra grande, estamos libres del sesgo de confusión.

En el resto de estrategias, partimos de un diseño de la intervención donde no se conoce completamente la ecuación estructural de participación en el tratamiento. Por un lado, para no tener estimaciones confundidas, Bell *et al.* (1995) proponen utilizar para el contrafactual a aquellos que solicitaron participar en el tratamiento y que finalmente no lo hicieron por una exclusión del implementador (basada solo en variables observadas). Controlando por dichas variables de exclusión o asumiendo que son independientes del resultado de interés nos libramos del sesgo de confusión. Por otro lado, la práctica común no es utilizar estos grupos de control internos, sino grupos de control externos (personas que ni siquiera solicitaron participar). Esto implica afirmar la asunción de no-confusión basada en las variables ajustadas. No obstante, tal como muestran Smith y Todd (2005), todavía no conocemos con certeza qué variables tendríamos que incluir en las estimaciones para reducir sustancialmente dicho sesgo. Por tanto, consideramos que una estimación tenía *mayor* validez interna si se basa en una de las estrategias sombreadas de la tabla 2, y que tenía *menor* validez interna en caso contrario³. Conformada la submuestra de mejor evidencia, realizamos una síntesis de sus resultados para aproximar el efecto medio de la ABE sobre las variables de interés.

TABLA 2. ESTRATEGIAS DE IDENTIFICACIÓN SEGÚN EL CONOCIMIENTO Y USO DE LA ECUACIÓN ESTRUCTURAL DE PARTICIPACIÓN EN EL TRATAMIENTO

		Usa para estimar el contrafactual a individuos que deciden participar pero que son excluidos por variables observadas	
		Sí	No
Conoce la ecuación estructural de participación en el tratamiento/control⁴	Sí	<ul style="list-style-type: none"> • Experimento con consentimiento. • Regresión discontinua con consentimiento. 	<ul style="list-style-type: none"> • Experimento sin consentimiento. • Regresión discontinua sin consentimiento.
	No	<ul style="list-style-type: none"> • Selección bajo observadas a la Bell <i>et al.</i> (1995). • DID a la Bell <i>et al.</i> (1995). 	<ul style="list-style-type: none"> • Selección bajo observadas clásica. • Regresión con variable instrumental. • DID clásica.

Fuente: elaboración propia. DID: diferencia en diferencias.

4. RESULTADOS

Una vez completada la búsqueda, se encontraron un total de veintidós estudios, que conformaron una muestra de 1042 estimaciones de impacto. La tabla 3 muestra la lista de estudios considerados en nuestro trabajo. Ninguno de ellos estuvo incluido en la revisión de Liu *et al.* (2014). Como podemos observar, la distribución de estimaciones de impacto por estudio es heterogénea y presenta notables casos atípicos. Es sobresaliente también el incremento de evaluaciones en los últimos años, habiéndose realizado solo seis evaluaciones antes del año 2010.

³ En realidad, las dos estrategias de la celda inferior izquierda podrían resumirse como «selección bajo observadas», ya que en realidad el ajuste fundamental es el ajuste por la variable *solicitó ser tratado*. Una vez que se controla esta variable, no es necesario controlar por sus antecesores, sean estos observados o no (Pearl, 2009).

⁴ Considérese un tratamiento binario voluntario. En realidad, hay tres grupos: tratamiento, control y no-participación en el estudio. La variable de las filas de la tabla 2 hace referencia a si podemos conocer la ecuación estructural de participación en el grupo tratamiento o de control, no a la ecuación estructural de participación en el grupo de no-participación en el estudio.

TABLA 3. MUESTRA DE ESTUDIOS INCLUIDOS

Estudio	Estimaciones	Territorio	Implementación
González Álvarez <i>et al.</i> (2021)	160	Aragón	[2015T1, 2018T4]
Di Paolo y Ramos (2020)	3	Cataluña	[2016T1, 2017T4]
Blazquez <i>et al.</i> (2019)	12	C. de Madrid	[2010T1, 2010T4]
CDI Consulting (2019)	5	Islas Canarias	[2018T1, 2018T4]
Company Alcover <i>et al.</i> (2019)	32	Islas Baleares	[2016T1, 2016T4]
AIReF (2018)	21	Aragón	[2016T1, 2016T4]
García-Pérez (2018)	21	España	[2011T3, 2015T2]
Notus (2018a)	344	C. de Madrid	[2016T1, 2017T4]
Notus (2018b)	84	C. de Madrid	[2016T1, 2017T4]
Di Paolo <i>et al.</i> (2017)	3	Cataluña	[2015T1, 2015T4]
Everis (2017)	5	Cataluña	[2016T1, 2016T4]
Cantero y Quintela (2016)	2	A Coruña	[2011T4, 2012T3]
Redcrea (2016)	2	España	[2015T1, 2015T2]
Alegre <i>et al.</i> (2015)	31	Barcelona	[2013T1, 2014T2]
ESI (2015)	140	Aragón	[2010T1, 2014T4]
Hodzic <i>et al.</i> (2015)	2	Valencia	[2012T4, 2014T2]
ESI (2010)	84	Aragón	[2005T1, 2009T4]
Ramos <i>et al.</i> (2009)	2	Cataluña	[2005T1, 2005T4]
Herrarte y Sáez (2008)	20	España	[2001T2, 2002T1]
ADETRE (2007)	4	Aragón	[2002T1, 2004T4]
EDEI (2007)	2	Islas Canarias	[2007T1, 2007T4]
Herrarte y Sáez (2007)	80	España	[2001T1, 2003T4]

Fuente: elaboración propia.

4.1. Los programas evaluados

En esta sección respondemos las dos primeras preguntas de investigación caracterizando qué programas de ABE han sido evaluados y cómo. Para ello, calculamos diferentes estadísticos de resumen partiendo la muestra según la variable respuesta. Recuérdese que la submuestra de empleo ($n_1 = 790$) es notablemente más grande que la submuestra de empleo de larga duración ($n_2 = 252$). La primera y tercera columnas de la tabla 4 y la tabla 5 presentan, para cada submuestra, la frecuencia absoluta y relativa de las estimaciones de impacto en función de diferentes variables. En la segunda y cuarta columna, vemos cuántos estudios (en términos absolutos y relativos) incluyen al menos una estimación de impacto sobre la categoría marcada en la fila.

En el caso de la submuestra de empleo, en términos de estudios, los componentes más evaluados son el itinerario, el establecimiento de metas y la información sobre tendencias de los mercados laborales. No obstante, la categoría más frecuente es *indeterminado*. Esto es problemático, ya que esta etiqueta recopila estimaciones de impacto de intervenciones muy heterogéneas, ya sea mediante un diseño personalizado o mediante una evaluación conjunta de diferentes programas. Atendiendo a la validez externa de las estimaciones, que el itinerario sea el componente principal más evaluado no es una buena noticia. Debemos considerar que los itinerarios son programas altamente adaptados al participante y que englo-

ban un conjunto diferente de posibilidades según la comunidad autónoma en cuestión. La submuestra de contratos largos es mucho más pequeña y tiene una distribución más uniforme con relación al componente principal. Vemos cómo existe un enorme vacío de estudios sobre los efectos de la ABE en la calidad del empleo. En ambas submuestras falta investigación sobre los efectos de programas con componentes de mejora de la motivación, como las técnicas relacionadas con la autoeficacia, el apoyo social y el manejo del estrés.

TABLA 4. FRECUENCIAS DE VARIABLES SOBRE EL DISEÑO DEL PROGRAMA

	Y = Pr[empleo]		Y = Pr[empleo de LD]	
	Estimación	Estudio	Estimación	Estudio
<i>Componente principal</i>				
Diagnóstico	26 (3,29)	3 (14,29)	22 (8,73)	2 (33,33)
Establecimiento de metas	162 (20,51)	7 (33,33)	22 (8,73)	2 (33,33)
Info. tendencias mercados laborales	94 (11,9)	5 (23,81)	22 (8,73)	2 (33,33)
Info. movilidad laboral	22 (2,78)	2 (9,52)	22 (8,73)	2 (33,33)
Itinerario	116 (14,68)	10 (47,62)	77 (30,56)	3 (50)
Hab. de autopresentación	44 (5,57)	2 (9,52)	44 (17,46)	2 (33,33)
Hab. de búsqueda de empleo	22 (2,78)	2 (9,52)	22 (8,73)	2 (33,33)
Autopresentación y búsqueda	77 (9,75)	4 (19,05)	0 (0)	0 (0)
Gestión del estrés	2 (0,25)	1 (4,76)	0 (0)	0 (0)
Indeterminado	225 (28,48)	10 (47,62)	21 (8,33)	2 (33,33)
<i>Sexo</i>				
Mujeres	117 (14,81)	13 (61,9)	43 (17,06)	5 (83,33)
Hombres	79 (10)	9 (42,86)	28 (11,11)	3 (50)
Mixto	594 (75,19)	20 (95,24)	181 (71,83)	5 (83,33)
<i>Edad</i>				
Jóvenes	133 (16,84)	9 (42,86)	42 (16,67)	4 (66,67)
Mixto	657 (83,16)	19 (90,48)	210 (83,33)	6 (100)
<i>Desaventajado</i>				
Desaventajado	200 (25,32)	8 (38,1)	156 (61,9)	4 (66,67)
Mixto	590 (74,68)	19 (90,48)	96 (38,1)	4 (66,67)
<i>Total de observaciones</i>	790	21	252	6

Nota: las entradas de celda son frecuencias absolutas, con las frecuencias relativas por columna entre paréntesis. En el caso de las columnas de estudios, pueden no sumar 22 ya que un mismo estudio puede contener estimaciones de diferentes categorías.

Fuente: elaboración propia.

Los efectos pueden evaluarse sobre diferentes grupos sociales, lo que depende tanto de los criterios de elegibilidad establecidos por el diseño del programa como del interés del investigador o del ente que encarga la evaluación. La tabla 4 muestra que, en la submuestra de empleo, la gran mayoría de las estimaciones analizan el efecto sobre un grupo mixto respecto al sexo, la edad o la desventaja. La estimación de efectos heterogéneos para las mujeres se realiza en la mayoría de los estudios, siendo menos

frecuentes las estimaciones para grupos desaventajados. No obstante, la proporción de estimaciones mixtas respecto a estimaciones para grupos concretos es muy alta. Solo un 17 % de las estimaciones se dedicaron a los jóvenes y solo una de cada cuatro se ocupó de las personas con desventajas. En el caso de la submuestra sobre empleo prolongado, la situación está más igualada. Es razonable que así sea teniendo en cuenta la mayor prevalencia de empleo precario en la población desaventajada, joven o mujer. No obstante, la prevalencia general de efectos agregados es preocupante debido a la evidencia presentada en metaanálisis previos como el de Card *et al.* (2018), donde se recogen efectos heterogéneos de las PAE según el grupo social.

TABLA 5. FRECUENCIAS DE VARIABLES METODOLÓGICAS

	$Y = Pr[\text{empleo}]$		$Y = Pr[\text{empleo de LD}]$	
	Estimación	Estudio	Estimación	Estudio
<i>Estrategia de identificación</i>				
Diferencia-en-diferencias	2 (0,25)	1 (4,76)	0 (0)	0 (0)
Experimento	3 (0,38)	2 (9,52)	1 (0,4)	1 (16,67)
Variable instrumental	22 (2,78)	1 (4,76)	9 (3,57)	1 (16,67)
Selección bajo observadas	763 (96,58)	17 (80,95)	242 (96,03)	4 (66,67)
<i>Tamaño de muestra*</i>				
Pequeña	36 (4,56)	7 (33,33)	25 (9,92)	3 (50)
Intermedia	143 (18,1)	9 (42,86)	117 (46,43)	4 (66,67)
Grande	122 (15,44)	6 (28,57)	100 (39,68)	3 (50)
<i>Horizonte de la VR</i>				
Muy corto plazo	335 (42,41)	16 (76,19)	118 (46,83)	4 (66,67)
Corto plazo	344 (43,54)	13 (61,9)	120 (47,62)	6 (100)
Medio plazo	109 (13,8)	4 (19,05)	14 (5,56)	2 (33,33)
Largo plazo	2 (0,25)	1 (4,76)	0 (0)	0 (0)
<i>Dimensión de la VR</i>				
Flujo	657 (83,16)	14 (66,67)	252 (100)	6 (100)
Stock	133 (16,84)	8 (38,1)	0 (0)	0 (0)
<i>Estimando</i>				
RD	710 (89,87)	20 (95,24)	252 (100)	6 (100)
OR, logit	80 (10,13)	3 (14,29)	12 (4,76)	1 (16,67)
<i>Inferencia = Sí</i>	102 (12,91)	12 (57,14)	38 (15,08)	4 (66,67)
<i>Publicado = Sí</i>	3 (0,38)	2 (9,52)	13 (5,16)	2 (33,33)

Nota: las entradas de las celdas son frecuencias absolutas (con las frecuencias relativas entre paréntesis, calculadas con el total de observaciones disponibles para cada variable). * El resto de las observaciones son valores no disponibles. LD: larga duración. VR: variable respuesta. RD: *risk difference*, también llamado efecto marginal. OR: *odds ratio*.

Fuente: elaboración propia.

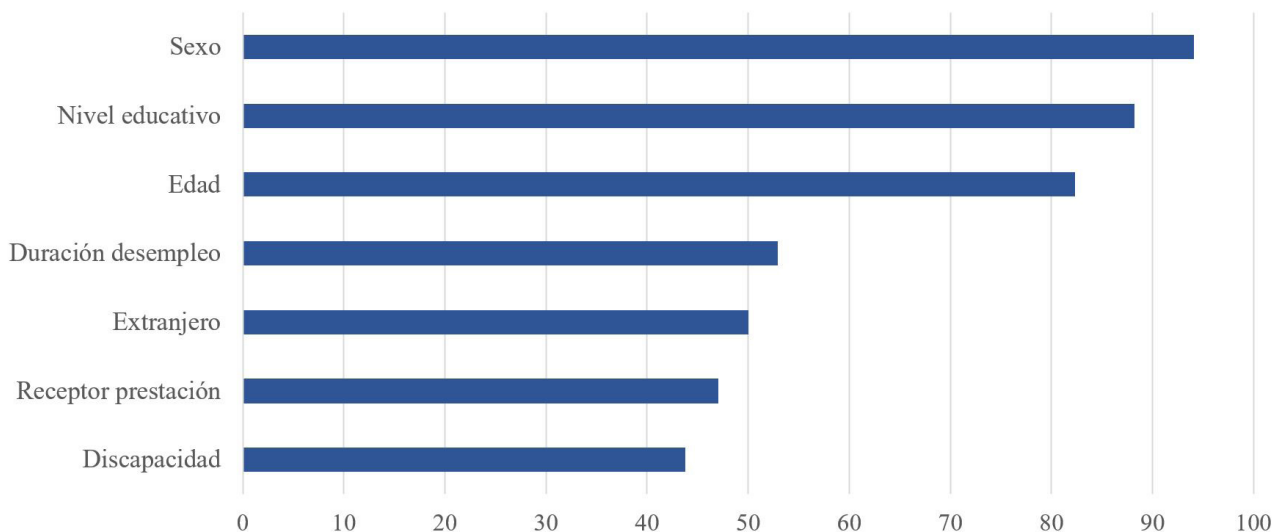
En cuanto a las decisiones metodológicas tomadas por los investigadores, como se muestra en la tabla 5, casi todos los estudios se han basado en una estrategia de selección bajo observadas. En la submuestra de

empleo solo hay cuatro estudios que emplearon una estrategia de identificación diferente. Tal como adelantamos en la sección metodológica, la selección bajo observadas es una estrategia difícil de justificar en cuanto ha de seleccionar un grupo de control sin saber con certeza qué variables de control incluir. Incluso en términos de comparación, en el gráfico 2 vemos cómo en la submuestra de empleo que usa esta estrategia solo hay tres variables ajustadas en más de dos tercios de los estudios: sexo, nivel educativo y edad. En la submuestra de empleo de larga duración, la distribución de estrategias de identificación es en general equivalente.

Mirando el tamaño de muestra, lo primero que notamos es que en la muestra de empleo el 61,9% de las estimaciones no reporta el número de observaciones en el que se basa. Esto está conectado con el hecho de que en ambas submuestras solo un poco más de la mitad de los estudios realizó algún procedimiento de inferencia. Pese a que el estudio trabaje con datos de toda la población de interés, creemos que es positivo realizar contrastes de hipótesis, tanto para descartar que los resultados sean fruto del azar como para poder generalizar los resultados a una superpoblación mayor. Siguiendo con la idea de generalizar los resultados, tal como han estudiado Colnet *et al.* (2023), no todos los estimandos permiten separar el efecto del tratamiento de la probabilidad que tiene el individuo de acceder al fenómeno de interés si no recibe el tratamiento (la línea de base). En nuestro caso, casi todas las observaciones son estimaciones de una diferencia de riesgos (RD). Tal como muestran Colnet *et al.* (í.d.), cuando la variable respuesta es binaria, la estimación del RD no aísla el efecto del tratamiento de las particularidades de nuestra muestra. Pese a ser práctica habitual, esto dificulta la validez externa de la evaluación, pues requiere de análisis futuros para transportar los resultados a otra población de interés.

En cuanto a la definición de la variable respuesta, encontramos que en ambas submuestras casi todas las estimaciones definieron el horizonte del efecto en el muy corto o corto plazo. Esto podría ser complementado en futuras evaluaciones, ya que tanto Card *et al.* (2018) como Kluve *et al.* (2019) encontraron mayores efectos en el largo plazo. En la submuestra de empleo, es preocupante la división de estudios según midan la variable respuesta en un solo punto temporal (dos tercios de los estudios) o en todo el intervalo temporal (algo más de un tercio). Esta heterogeneidad dificulta la comparación de resultados entre estudios, ya que en realidad están midiendo cosas diferentes. Por el contrario, la submuestra de empleo prolongado solo presenta variables dependientes definidas como flujos.

GRÁFICO 2. PROPORCIÓN DE ESTUDIOS QUE AJUSTARON POR CADA UNA DE LAS SIGUIENTES VARIABLES

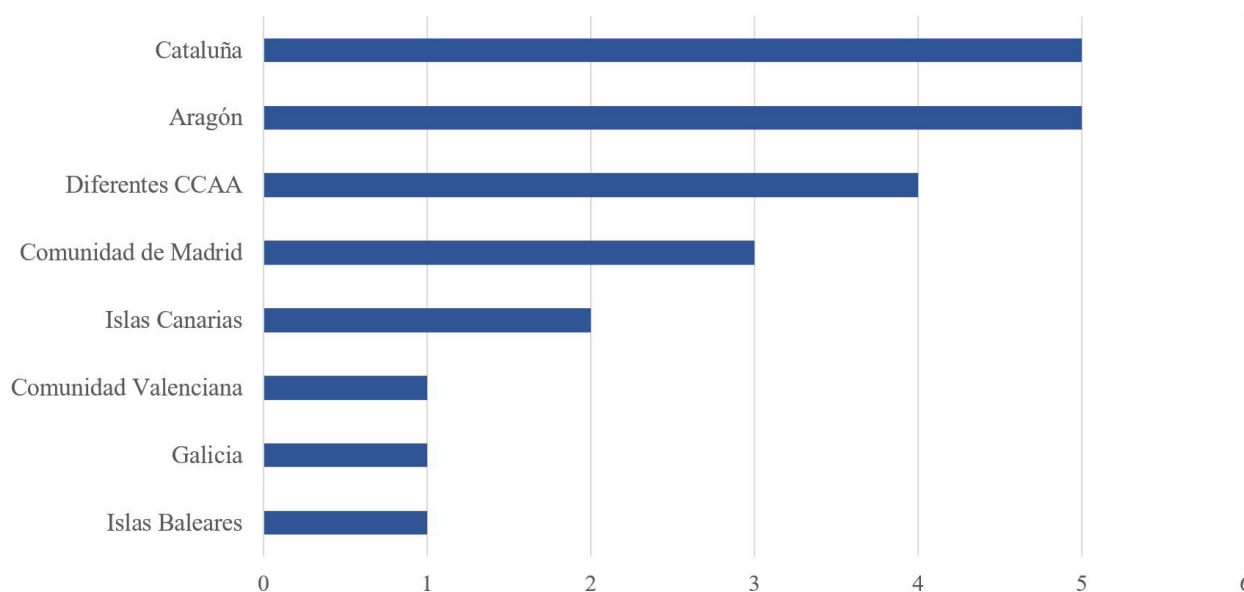


Nota: solo se incluyen variables ajustadas en al menos un tercio de los estudios y solo se consideran los estudios que usaron la selección bajo observadas como estrategia de identificación y Pr[empleo] como variable respuesta.

Fuente: elaboración propia.

Respecto al contexto de implementación de los programas evaluados, en primer lugar, el gráfico 3 nos muestra la distribución de estudios por comunidad autónoma. Como vemos, hay una notable diferencia entre

las dos primeras comunidades y el resto. En el caso de Aragón y Cataluña, tenemos cinco estudios centrados en cada una de ellas. Si miramos los años de publicación y la literatura relacionada, vemos que estas autonomías tienen una mayor tradición evaluadora. Esto contrasta con comunidades menos desarrolladas, como Andalucía o Extremadura, que no presentan ningún estudio que se haya ocupado exclusivamente de programas implementados en estos territorios. Contamos con algunas estimaciones sobre estas comunidades, concretamente cinco, sobre la primera y tres sobre la segunda, pero hacen referencia a programas implementados hace más de dos décadas. Tal como observamos en la tabla 3, el periodo de implementación de las estimaciones analizadas se sitúa fundamentalmente en la década de los 2010. Esto nos indica el considerable aumento de la evidencia disponible en los últimos años. Por último, con la muestra a nivel de estudio, el promedio de las tasas de paro que tuvieron lugar durante los periodos analizados es de un 15,54 %, con una desviación típica de un 5,58 %. Por tanto, tenemos evaluaciones tanto de momentos de bonanza como de recesión en el mercado laboral, lo que dificulta la comparabilidad de resultados si existiera interacción.

GRÁFICO 3. NÚMERO DE ESTUDIOS POR COMUNIDAD AUTÓNOMA

Fuente: elaboración propia.

4.2. Los efectos de los programas

En este apartado realizaremos un análisis detallado de la mejor evidencia disponible. La tabla 6 recoge aquellos artículos que han pasado el filtro metodológico que establecimos en relación con la validez interna. En la literatura actual tenemos cuatro estudios con una metodología suficientemente robusta al sesgo de confusión, que se concretan en once estimaciones de impacto de intervenciones de ABE. Por un lado, tanto Hodzic *et al.* (2015) como Cantero y Quintela (2016) realizaron un experimento aleatorizado. Esto implica que, en muestras grandes, la estimación del efecto estaría libre de confusión por el propio diseño de la intervención. Por otro lado, Redcrea (2016) y Everis (2017) aplicaron estrategias de identificación donde el grupo de control se escogió entre individuos en lista de espera que habían solicitado participar, pero que fueron descartados por los implementadores del programa. En ambos casos, la exclusión del programa estuvo motivada por el limitado número de plazas disponibles. En estos estudios, la estimación del efecto estaría libre de confusión al ajustar por la variable de solicitud de participación.

Mirando individualmente los resultados de cada estudio, es difícil llegar a conclusiones claras. La evaluación de Hodzic *et al.* (2015) solo presenta dos estimaciones. Aunque el efecto de la intervención a los seis meses parece nulo (positivo, pero no significativo), un año después del tratamiento el impacto se vuelve positivo. Además, la estimación puntual de dicha ganancia de probabilidad (26,4 %) es considerablemente alta. En comparación con el resto de estimaciones de mejor evidencia disponible, las de Hodzic *et al.* (id.) han de tomarse con precaución, ya que mide la variable respuesta en formato stock. Esto refuerza todavía

más lo positivo de sus resultados. La otra evaluación experimental, la de Cantero y Quintela (2016), no encontró un efecto sobre el empleo en el corto plazo. A pesar de su pequeña muestra, podríamos descartar que este resultado estadísticamente no significativo se explicara por una baja potencia estadística, ya que la estimación puntual está muy cercana a cero (concretamente es el 0,68 %). Por último, este estudio sí que aproxima el efecto sobre la probabilidad de conseguir un empleo prolongado, obteniendo una estimación del 13,6 %. Un coeficiente de una magnitud notable si lo comparamos con la mayoría de estimaciones del efecto sobre la probabilidad de conseguir empleo.

Los resultados de los estudios observacionales también presentan una imagen inconcluyente sobre los efectos de la ABE. En el caso de Everis (2017), las cinco estimaciones son acerca del muy corto plazo y tienen coeficientes aproximadamente igual a cero. Dicha ausencia de efectos contrasta con el incremento de un 18,4 % en la probabilidad de encontrar empleo que encuentran para el programa Lanzaderas de Empleo los autores de Redcrea (2016).

TABLA 6. MEJOR EVIDENCIA DISPONIBLE SOBRE ABE PARA ESPAÑA

Estudio	Efectos estimados	N.º de estimaciones y variable respuesta	Estrategia de identificación
Everis (2017)	E: [- 0,03, 0]	5 x MCP-f	Selección bajo observadas à la Bell <i>et al.</i> (1995)
Redcrea (2016)	E: [0,034, 0,184]	1 x MCP-s 1 x MCP-f	DID à la Bell <i>et al.</i> (1995)
Cantero y Quintela (2016)	E: [0,007]	1 x CP-f	Experimento con consentimiento
	EL: [0,136]	1 x CP-f	
Hodzic <i>et al.</i> (2015)	E: [0,112, 0,264]	1 x MCP-s 1 x CP-s	Experimento con consentimiento

E: efecto sobre la Pr[empleo]. EL: efecto sobre la Pr[empleo de larga duración]. En ambos casos, medido como diferencia de riesgos. La columna de efectos estimados indica el rango de los efectos estimados. MCP: muy corto plazo, CP: corto plazo, -s: dimensión «stock», -f: dimensión «flujo».

Fuente: elaboración propia.

Si queremos resumir toda la información disponible, centrándonos en los efectos sobre la probabilidad de encontrar empleo a muy corto plazo, el efecto promedio es de un 2,71 %. Esta estimación ha de tomarse con cautela, ya que incluye tanto estimaciones con variable respuesta de tipo stock como de tipo flujo. El siguiente paso es profundizar en los detalles de diseño, metodología y contexto de cada una de las evaluaciones. Los diferentes efectos que se presentan a continuación, condicionando por algunas de estas variables, han de interpretarse como indicadores de asociación, no como consecuencias del diseño, el método o el contexto. El énfasis en uno u otro dependerá de las asunciones acerca de cuáles son las variables modificadoras de los efectos de la ABE.

Respecto al diseño de los programas, solo la intervención de Hodzic *et al.* (2015) presenta claramente cuál es su componente principal, siendo este la gestión del estrés. La aparición de un efecto mayor a corto plazo en Hodzic *et al.* (ídem.) es consistente con la teoría autorregulatoria, ya que la gestión del estrés incidiría principalmente en el último proceso (el de análisis). El resto de estudios presentan sus intervenciones como un conjunto de acciones de búsqueda de empleo con diferentes componentes, donde en algunos casos tal conjunto parece variar entre individuos (Cantero y Quintela, 2016; Everis, 2017), mientras que en otros casos no parece variar (Redcrea, 2016; Hodzic *et al.*, 2015). En el caso de Cantero y Quintela (2016), dicho conjunto puede incluso contener acciones de formación de adultos. Condicionando por esta variable, podemos ver cómo aquellas intervenciones donde la secuencia de acciones estuvo determinada por el implementador tienen de media efectos estimados sobre el empleo más altos (0,148 vs. -0,022).

Si nos fijamos en las características de los grupos diana analizados en cada una de las estimaciones, veremos que no existen grandes diferencias en los efectos promedios sobre la probabilidad de empleo. Si condicionamos por la edad, no aparecen diferencias sustantivas entre aquellas estimaciones de efecto que

consideraron grupos mixtos en edad (0,051) frente a aquellas que consideraron grupos de jóvenes (0,04). Tampoco parecen ser diferentes los efectos por sexo, ya que el efecto promedio para las estimaciones con grupos de solo mujeres (-0,012) son prácticamente iguales a las de solo hombres (-0,02). No podemos saber si los grupos desaventajados se benefician en mayor medida de estos programas porque la submuestra de mejor evidencia disponible solo incluye estimaciones mixtas respecto a esta variable.

Con idea de trasladar lo aprendido en estas evaluaciones a futuras intervenciones, es preciso analizar la validez externa de estos estudios. Para ello, es importante considerar tanto variables metodológicas como de contexto. En primer lugar, para facilitar la transportabilidad, es importante que el tamaño de muestra sea suficientemente grande para que los soportes de las covariables moderadoras sean lo más amplios posibles (Colnet *et al.*, 2023). Tres de los cuatro estudios seleccionados manejan muestras que hemos clasificado como pequeñas. La excepción, Everis (2017), utiliza muestras de tamaño intermedio. El uso de muestras grandes también contribuye a incrementar la validez interna, ya que los contrastes de hipótesis más usados en economía están basados en teoría asintótica. Nótese que los dos estudios observacionales no realizaron ningún procedimiento de inferencia. Así, focalizándonos en resultados que sabemos que no son fruto del azar, tendríamos un efecto nulo al muy corto plazo, pero positivo a corto plazo.

Si consideramos el contexto de implementación como un posible factor moderador del efecto de la ABE, la evaluación de Redcrea (2016) toma especial fuerza. Este estudio analiza un programa implementado en diez municipios de siete comunidades autónomas diferentes, casi todos ellos de un tamaño intermedio. Esto garantiza una mayor pluralidad de mercados laborales locales frente a otras evaluaciones centradas en grandes ciudades como Barcelona (Everis, 2017) o Valencia (Hodzic *et al.*, 2015). No obstante, también hemos de considerar que las cuatro evaluaciones incluidas en esta sección realizaron sus estimaciones en periodos de alto desempleo, con tasas de paro autonómicas que van desde el 17,96 % (Everis, 2017) hasta el 27,18 % (Hodzic *et al.*, 2015). Si consideráramos factible que esta variable sea una moderadora del efecto de la ABE, necesitaríamos más estudios que analicen el efecto de la intervención en periodos de menor tasa de paro.

5. DISCUSIÓN

5.1. Sobre la variable de tratamiento

Una de las deficiencias más generalizadas de los estudios revisados ha sido la falta de una descripción detallada del diseño del programa. Podría ser útil que futuras evaluaciones incorporaran una sección que explique la intervención tal como se planeó en el diseño del programa y/o tal como se implementó. Esto permitiría distinguir adecuadamente las diferentes intervenciones. Schneider e Ingram (1990) propusieron una lista de seis elementos que caracterizarían el diseño de toda política: metas, agentes, poblaciones objetivo, reglas, herramientas y supuestos. En términos más prácticos, esto supondría incluir, al menos, la teoría del cambio, el número de sesiones de cada tratamiento (si no es fijo, la media y desviación estándar), la duración de cada sesión, el mecanismo de asignación de cada sesión (en caso de programas de duración variable) y los componentes de cada sesión. Dicha descripción puede obtenerse a través del análisis de datos cualitativos, como textos normativos (órdenes, resoluciones...) o entrevistas a informantes clave (diseñadores, proveedores, participantes...).

En intervenciones hechas a medida (como los itinerarios), la situación ideal sería evaluar por separado diferentes combinaciones de acciones. Para que la evidencia empírica sea adecuada para la inclusión en una futura revisión de literatura, debemos pensar que programas con el mismo nombre no siempre incluyen las mismas acciones en diferentes comunidades autónomas. Por ejemplo, en nuestra muestra los itinerarios en Canarias pueden incluir formación de adultos, mientras que en las Islas Baleares pueden estar relacionados con subvenciones al empleo público o privado. Por tanto, es útil reportar tanto información sobre el diseño como sobre la implementación.

5.2. Sobre la variable respuesta

También hemos encontrado una considerable heterogeneidad en la forma de medir la variable respuesta. En primer lugar, respecto a *quién* se mide no existe consenso sobre la definición de grupos sociales. Esto hace que algunos estudios fijen como submuestra de jóvenes a los menores de treinta años, mientras

que otros consideran los treinta y cinco como umbral. Es también el caso del colectivo de parados de larga duración. Algunos autores o Gobiernos consideraban a una persona como tal si permanecía más de doce meses desempleada, mientras que otros establecían dieciocho meses como mínimo. La heterogeneidad es aún mayor en el superconjunto de personas en riesgo de exclusión. En estos casos, la coordinación entre los evaluadores sería fructífera para producir evidencia empírica lo suficientemente comparable para futuras revisiones de literatura. Nuestra sugerencia es que los investigadores proporcionen estadísticos descriptivos detallados sobre la composición sociodemográfica de los grupos para permitir la transportabilidad de sus estimaciones.

En segundo lugar, sobre *cuándo* se mide el efecto, creemos que la opción de flujo ofrece una observación más completa de los efectos y es factible al tener acceso a registros administrativos. Además, hay otros tres aspectos del horizonte de efectos que merecerían más atención. Por un lado, sería interesante seguir definiciones iguales de muy corto, corto, medio y largo plazo. Aquí hemos seguido uno de los trabajos más citados en la literatura (Card *et al.*, 2018), que propone doce, veinticuatro y treinta y seis meses para los tres últimos horizontes. Por otro lado, los investigadores podrían aclarar en mayor medida si están analizando efectos *ex-ante*, *ex-dure* y/o *ex-post*. Algunos trabajos como García-Pérez (2018) ya han arrojado luz sobre este punto. Sugeriríamos, al menos, adoptar su enfoque e incluso reconocer la posibilidad de efectos anticipatorios. Estimar exclusivamente efectos a muy corto plazo podría ofrecer al diseñador una visión parcial del impacto del programa. Los efectos pueden desvanecerse o, incluso, aparecer más adelante en función de los mecanismos implicados. Sabemos que esto último es común en programas de formación.

Por último, sería útil explicitar si hemos establecido para la medición de los efectos horizontes individuales o un horizonte común. Intervenciones como la ABE suelen presentar diferentes fechas de inicio y finalización en función del participante, ya que no suelen ser intervenciones grupales. Un horizonte individual consideraría la fecha final del tratamiento de cada individuo en concreto y observaría los resultados durante los seis siguientes meses. Esta sería nuestra opción preferida. Un horizonte común establecería los mismos seis meses de observación para cada participante, independientemente del mes en que terminó el tratamiento. Este horizonte común podría elegirse atendiendo a la fecha de finalización mediana o media.

5.3. Sobre la estrategia analítica

Cuando el objetivo de un estudio es conocer el efecto de una intervención, es común considerar dos propiedades del análisis: la validez interna y la validez externa (Cook y Campbell, 1979). La validez interna de un estudio suele juzgarse a partir de su estrategia de identificación, es decir, a partir del conjunto de asunciones por las que sustituimos el término contrafactual (no observado) por una cantidad observada. Como hemos visto, la literatura sobre los efectos de la ABE en España se ha basado principalmente en una estrategia de identificación: la selección bajo observadas. Esta estrategia presenta el problema de la selección arbitraria de variables para el procedimiento de ajuste. Las implicaciones que puede tener la formulación de una política pública en base a una estimación confundida del efecto pueden ser graves. Nótese que el sesgo de variable omitida puede incluso cambiar la dirección del efecto estimado. Por ello, usar evidencia proveniente de estrategias con grandes asunciones, como la selección bajo observadas, puede dar la falsa impresión de estar utilizando resultados robustos sin el coste de tener que realizar un experimento. Esto solo es cierto bajo condiciones muy estrictas que han de evaluarse para cada caso y que desconocemos si aplican actualmente. Si la realización de un experimento aleatorizado no es posible, nuestra propuesta es utilizar las listas de espera para formar grupos de control tal como proponen Bell *et al.* (1995). Cuando la variable que determina la inclusión en la lista de espera es independiente del resultado de interés⁵, los análisis podrían ir incluso más allá de lo que permite una estrategia de regresión discontinua basada en una variable de puntuación relacionada con la variable resultado. Nótese que la estrategia de grupos de control internos ya se ha explorado en España con intervenciones de formación (Cueto y Mato, 2009).

Una vez que aplicamos una estrategia de identificación creíble, hay procedimientos que ayudan a interpretar la magnitud del efecto. Pese a que la población de interés sea una población finita, nuestra propuesta es realizar contrastes de hipótesis para facilitar la interpretación de efectos muy pequeños. De este modo,

⁵ Podría ser, por ejemplo, el orden de llegada.

pensando también en futuras revisiones de literatura, proponemos que siempre se incluyan los estadísticos de los procedimientos de inferencia (p-valor, estadístico de contraste, error estándar...) y los tamaños de muestra. Podríamos fijarnos en iniciativas como TRIPOD o STROBE, desarrolladas por investigadores de la salud. Son listas de verificación sobre cada elemento que un estudio debería incorporar para ser analizado posteriormente en un metaanálisis.

La validez externa de un análisis se preocupa por cuánto de generalizables son nuestros resultados. En primer lugar, deberíamos considerar que la selección de un estimando u otro (por ejemplo, una diferencia de riesgos en vez de una ratio de riesgos) impacta sobre la facilidad con la que se pueden transportar los resultados (Colnet *et al.*, 2023). En segundo lugar, la llamada a realizar más análisis de efectos a medio y largo plazo ha de considerar si estamos interesados en estimar el efecto total o el efecto directo de una intervención (Frölich y Sperlich, 2019). El efecto total incluye los efectos directos de otras intervenciones cuya participación fue causada por la ABE; por ejemplo, un curso de formación. El efecto directo excluye los efectos de estas variables mediadoras. Hemos de tener en cuenta que la estrategia de identificación de estos últimos es generalmente más exigente que la del efecto total, que es lo que nuestros estudios han identificado implícitamente. Una solución intermedia podría ser o bien estimar los efectos de la ABE sobre la participación en otras intervenciones o bien simplemente reportar estadísticos de resumen sobre la participación en dichas intervenciones posteriores.

6. CONCLUSIÓN

En este estudio hemos realizado una revisión de literatura acerca de los trabajos que han estimado efectos sobre el empleo y sobre el empleo prolongado de la ABE en España. Para ello, hemos construido una muestra de 1042 observaciones a partir de veintidós estudios. Dicha muestra ha sido analizada de manera exhaustiva y, posteriormente, seleccionando las evaluaciones con mayor validez interna.

La primera pregunta trató sobre qué programas de ABE se han evaluado. En cuanto al componente principal, los itinerarios son los programas que han sido evaluados más veces. Sin embargo, hemos detectado en muchos casos tanto una falta de detalle a la hora de describir los diseños como la agregación de diversas intervenciones en un solo tratamiento, lo que complica la transportabilidad de resultados. En cuanto a los territorios evaluados, hemos visto que existen comunidades autónomas con una evaluación frecuente de la ABE (Cataluña y Aragón), mientras que otras todavía no han realizado ninguna evaluación de impacto sobre las acciones que financia (p.ej., Andalucía y Extremadura).

La segunda pregunta planteó cómo se han evaluado dichas intervenciones. El primer resultado es que el 80% de estudios que estiman el efecto sobre el empleo utilizan una identificación por selección bajo observadas, siendo esta la estrategia con supuestos más fuertes. Esto hace menos creíble que sus resultados sean verdaderamente capaces de aislar el impacto de la intervención. El segundo resultado es la notable presencia de decisiones metodológicas que dificultan la generalización de resultados: diferente definición de la variable respuesta, no reporte de pruebas de inferencia y no utilización de *estimandos* más transportables como la ratio de riesgos. La implicación de esto es que necesitamos un mayor acuerdo entre la comunidad evaluadora.

La tercera pregunta pretendió estimar el efecto promedio de la ABE sobre el empleo y el empleo prolongado. Centrándonos en los estudios con alta validez interna, hemos visto que aquellas intervenciones que tuvieron una secuencia de acciones determinada por el implementador presentaron un mayor efecto sobre el empleo (14,8% vs. -0,02%). No obstante, si consideramos todas las estimaciones del efecto sobre el empleo de esta submuestra, los resultados son ambiguos, ya que el efecto promedio es de tan solo un 2,71%, partiendo además de un número pequeño de observaciones.

Por último, el análisis de la literatura indica algunas líneas de investigación futura que nos gustaría destacar. Respecto al diseño, proponemos acompañar las evaluaciones de impacto de evaluaciones de diseño e implementación para poder definir correctamente el tratamiento. Respecto al impacto, los resultados de la primera pregunta de investigación señalan tres grandes huecos donde precisamos más evaluación de impacto: efectos sobre el empleo prolongado, de intervenciones motivacionales y sobre horizontes a medio/largo plazo. Los datos administrativos actuales incluyen información sobre algunas de estas variables que pueden ser útiles. Algunos de los estudios revisados nos han mostrado que es posible estimar los efectos con una mayor validez interna usando estos datos con estrategias de identificación más sólidas.

AGRADECIMIENTOS

Agradecemos a los evaluadores anónimos y al equipo editorial de la revista sus sugerencias y comentarios a la primera versión de este artículo.

FINANCIACIÓN

Investigación financiada por el Ministerio de Universidades del Gobierno de España en el marco del Plan de Formación del Profesorado Universitario (FPU) con referencia FPU19/05291.

REFERENCIAS BIBLIOGRÁFICAS

- ADETRE (2007). *Valoración del impacto económico y social de las políticas activas del Instituto Aragonés de Empleo*. INAEM.
- Alegre, M. Á., Todeschini, F. y Rodés, M. (2015). *Avaluació de l'impacte dels Serveis d'Orientació i Recerca de Feina de Barcelona Activa*. Ivàlua.
- AIReF (2018). *Evaluación del gasto público 2018. Proyecto 3 (PAE). Estudio programa políticas activas de empleo*. AIReF.
- Arranz, J. M. y García Serrano, C. (2022). La evaluación del impacto de las políticas públicas: el caso de las políticas del mercado de trabajo. *Papeles de Economía española*, (172), 139-154.
- Bell, S. H., Orr, L. L., Blomquist, J. D. y Cain, G. G. (1995). *Program applicants as a comparison group in evaluating training programs: Theory and a test*. W. E. Upjohn Institute for Employment Research. <https://doi.org/10.17848/9780585284545>
- Blazquez, M., Herrarte, A. y Saez, 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>
- Bonoli, G. (2013). *The origins of active social policy: Labour market and childcare policies in a comparative perspective*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199669769.001.0001>
- Brown, A. J., y Koettl, J. (2015). Active labor market programs—employment gain or fiscal drain? *IZA Journal of Labor Economics*, 4(1), 1-36. <https://doi.org/10.1186/s40172-015-0025-5>
- Cahuc, P., Carcillo, S. y Zylberberg, A. (2014). *Labor economics*. MIT press.
- Caliendo, M. y Künn, S. (2011). Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity. *Journal of Public Economics*, 95(3-4), 311-331. <https://doi.org/10.1016/j.jpubeco.2010.11.003>
- Cantero, J. M. M. y Quintela, N. R. (2016). Evaluación del impacto de políticas orientadas a la inserción sociolaboral mediante matching. *Revista de Investigación Educativa*, 34(1), 87-102. <https://doi.org/10.6018/rie.34.1.206871>
- Card, D., Kluve, J. y 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/jeaa/jvx028>
- Cárdenas del Rey, L. (2020). La segmentación laboral durante la recuperación económica: empleo atípico y rotación. *Cuadernos de Relaciones Laborales*, 38(1), 145. <https://doi.org/10.5209/crla.68871>
- Catalá-Pérez, D. y Del Pino, E. (2018). Reformas de ida y vuelta en la gobernanza horizontal del bienestar: el caso de la colaboración público-privada en la intermediación laboral en la Comunidad Valenciana. *Gestión y Análisis de Políticas Públicas*, 19, 107-122. <https://doi.org/10.24965/gapp.v0i19.10434>
- CDI Consulting (2019). *Evaluación de los objetivos/resultados del PO FSE 2014-2020 de Canarias para el informe anual de 2018 a remitir en 2019 y de cumplimiento del marco de rendimiento. Informe de evaluación final*. Dirección General de Planificación y Presupuesto, Gobierno de Canarias.
- Cheung, M., Egebark, J., Forslund, A., Laun, L., Rödin, M. y Vikström, J. (2023). Does Job Search Assistance Reduce Unemployment? Evidence on Displacement Effects and Mechanisms. *Journal of Labor Economics*. <https://doi.org/10.1086/726384>
- Colnet, B., Josse, J., Varoquaux, G. y Scornet, E. (2023). Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize? *arXiv*, arXiv:2303.16008v2. <https://doi.org/10.48550/arXiv.2303.16008>
- Cook, T. D., y Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin.
- Company Alcover, M., Massanet Isern, E., Pou Garcias, Ll. y Quetglas Oliver, M. (2019). Evaluación de las políticas activas de ocupación. El caso de la orientación laboral a personas de colectivos vulnerables. *Focus Econòmic*, 11. SOIB.
- Crépon, B., y Van Den Berg, G. J. (2016). Active labor market policies. *Annual Review of Economics*, (8), 521-546. <https://doi.org/10.1146/annurev-economics-080614-115738>
- Cueto, B., y Mato, F. J. (2009). A nonexperimental evaluation of training programmes: regional evidence for Spain. *The Annals of Regional Science*, (43), 415-433. <https://doi.org/10.1007/s00168-008-0214-2>

- Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press. <https://doi.org/10.2307/j.ctv1c29t27>
- Di Paolo, A. y Ramos, R. (2020). *Avaluació de la Implementació i de l'Impacte del Programa Noves Oportunitats a Catalunya, Edició 2015-2017*. SOC.
- Di Paolo, A., Ramos, R. y Suriñach, J. (2017). *Evaluación de la implementación del programa de Acciones de orientación y acompañamiento a la inserción de la red de orientación para el empleo que desarrolla el Servicio Público de Empleo de Catalunya*. SOC.
- DG EMPL (2023). *Labour Market Policy* (Versión 2023.07.14). Comisión Europea, Dirección General de Empleo, Asuntos Sociales e Inclusión.
- EDEI (2007). *Evaluación de las políticas públicas de empleo gestionadas por el SCE. Informe Final de Evaluación políticas gestionadas 2007*. SCE.
- Everis (2017). *Evaluación de implementación y de impacto de los Servicios de Orientación Profesional*. SOC.
- ESI (2010). *Valoración del impacto económico y social de las políticas activas del Instituto Aragonés de Empleo*. INAEM.
- ESI (2015). *Estudio sobre la valoración económica y social de las políticas activas del Instituto Aragonés de Empleo*. INAEM.
- Frölich, M. y Sperlich, S. (2019). *Impact Evaluation. Treatment Effects and Causal Analysis*. Cambridge University Press. <https://doi.org/10.1017/9781107337008>
- García-Pérez, J. I. (2018). Una primera evaluación del impacto sobre la salida del desempleo de las políticas activas ofrecidas por los servicios públicos de empleo en España. En F. Felgueroso, J. I. García-Pérez and S. Jiménez-Martín (coords.), *Perfilado estadístico*. Fundación Ramon Areces and FEDEA.
- González Álvarez, M. A., Ben Kaabia, M., Clemente López, J., García Castrillo, P., Montañés Bernal, A. y Olmos Salvador, L. (2021). *Estudio sobre la valoración del impacto económico y social de las Políticas Activas del Instituto Aragonés de Empleo: 2015-2018*. INAEM.
- Hernández-Moreno, J. J., y Ramos Gallarín, J. A. (2017). La centralización de la gobernanza territorial en las políticas activas del mercado de trabajo en España durante la Gran Recesión (2011-2016). *Gestión y Análisis de Políticas Públicas*, 18, 5-22. <https://doi.org/10.24965/gapp.v0i18.10433>
- Herrarte, A. y Sáez, F. (2007). *Labour Market Policy in Spain: Analysis of microdata and main results. Labour Market Policy Seminar 2007 Edition*. Eurostat and DG Employment, Social Affairs and Equal Opportunities.
- Herrarte, A. y Sáez, F. (2008). *An evaluation of ALMP: The case of Spain* [Paper No. 55387]. Munich Personal RePEc Archive – MPRA.
- Hodzic, S., Ripoll, P., Lira, E., y Zenasni, F. (2015). Can intervention in emotional competences increase employability prospects of unemployed adults? *Journal of Vocational Behavior*, (88), 28-37. <https://doi.org/10.1016/j.jvb.2015.02.007>
- Kluge, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., Weidenkaff, F. y Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, (114), 237-253.
- Liu, S., Huang, J. L., y Wang, M. (2014). Effectiveness of job search interventions: A meta-analytic review. *Psychological bulletin*, 140(4), 1009. <https://doi.org/10.1037/a0035923>
- Malo, M. Á. (2018). *Finding proactive features in labour market policies: A reflection based on the evidence* [Research Papers Series, 8]. International Labour Organization – ILO. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_650075.pdf
- Malo, M. Á. y Cueto, B. (2015). El impacto de las políticas activas de mercado de trabajo en España. *Documentación social*, (178), 105-120.
- Manski, C. F. (2007). *Identification for Prediction and Decision*. Harvard University Press. <https://doi.org/10.2307/j.ctv219kxm0>
- Marinescu, I. E. (2017). *Job search monitoring and assistance for the unemployed*. IZA World of Labor. <http://dx.doi.org/10.15185/izawol.380>
- Notus (2018a). *Evaluación del Impacto Individual de Políticas Activas de Empleo 2018. Todos los demandantes*. Dirección General del Servicio Público de Empleo de la Comunidad de Madrid.
- Notus (2018b). *Evaluación del Impacto Individual de Políticas Activas de Empleo 2018. Informe de colectivos*. Dirección General del Servicio Público de Empleo de la Comunidad de Madrid.
- Pearl, J. (2009). *Causality*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803161>
- Petticrew, M. y Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley and Sons. <https://doi.org/10.1002/9780470754887>
- Ramos, R., Suriñach, J. y Artís, M. (2009). *La efectividad de las políticas activas de mercado de trabajo para luchar contra el paro. La experiencia de Cataluña* [Working Papers, 2009/19]. Research Institute of Applied Economics.
- Ramos, R., Suriñach, J. y Artís, M. (2010). ¿Es necesario reformar las políticas activas de mercado de trabajo en España? Algunos elementos para la reflexión. *Papeles de economía española*, (124), 281-300.
- Redcrea (2016). *Evaluación del impacto social de las Lanzaderas de Empleo*. Fundación Telefónica y Fundación Santa María La Real.
- Schneider, A. L. y Ingram, H. M. (1990). Policy design: Elements, premises, and strategies. En S. Nagel (ed.), *Policy theory and policy evaluation*. Greenwood Press.

- Slavin, R. E. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 9-18. [https://doi.org/10.1016/0895-4356\(94\)00097-A](https://doi.org/10.1016/0895-4356(94)00097-A)
- Smith, J. A. y 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>
- Vooren, M., Haelermans, C., Groot, W. y Maassen van den Brink, H. (2019). The effectiveness of active labor market policies: a meta-analysis. *Journal of Economic Surveys*, 33(1), 125-149. <https://doi.org/10.1111/joes.12269>