

GAPP. Nueva época, número 31, marzo de 2023  
Monográfico: «Enseñar las políticas públicas»: el qué, el cómo y el dónde  
Sección: ARTÍCULOS  
Recibido: 04-10-2022  
Modificado: 09-02-2023  
Aceptado: 09-02-2023  
Prepublicado: 24-02-2023  
Publicado: 01-03-2023  
DOI: <https://doi.org/10.24965/gapp.11133>  
Páginas: 56-70



Referencia: Miller, L., Rodríguez Marín, I. (2023). Estrategias experimentales para el análisis y evaluación de políticas públicas. *Gestión y Análisis de Políticas Públicas*. 31, 56-70. <https://doi.org/10.24965/gapp.11133>

## Estrategias experimentales para el análisis y evaluación de políticas públicas<sup>1</sup>

### *Experimental strategies for public policy analysis*

Miller, Luis

Consejo Superior de Investigaciones Científicas – IPP-CSIC (España)

ORCID: <https://orcid.org/0000-0002-7447-2313>

[luis.miller@csic.es](mailto:luis.miller@csic.es)

#### NOTA BIOGRÁFICA

Científico titular del Instituto de Políticas y Bienes Públicos del CSIC. Trabajó como investigador posdoctoral en el Instituto Max Planck de Economía y como adjunto a la dirección e investigador posdoctoral del Centro de Ciencias Sociales Experimentales del Nuffield College de la Universidad de Oxford. Desde 2011 hasta 2018 fue profesor permanente de microeconomía y economía experimental y conductual en la Facultad de Economía y Empresa de la Universidad del País Vasco. Ha sido asesor en la Oficina Nacional de Prospectiva y Estrategia de la Presidencia del Gobierno de España, Director de Gabinete de la Presidencia del CSIC y actualmente preside el Comité ético vinculado a los itinerarios de inclusión social del Ministerio de Inclusión, Seguridad Social y Migraciones. Publica regularmente sobre justicia distributiva, polarización política y métodos experimentales en disciplinas como la economía, ciencia política, sociología, psicología y filosofía.

Rodríguez Marín, Isabel

Consejo Superior de Investigaciones Científicas – IPP-CSIC (España)

ORCID: <https://orcid.org/0000-0003-2363-3700>

[isabel.rodriquez@csic.es](mailto:isabel.rodriquez@csic.es)

#### NOTA BIOGRÁFICA

Investigadora predoctoral en el Instituto de Políticas y Bienes Públicos del CSIC, con un contrato FPU del Ministerio de Universidades. Doctoranda en el programa Interuniversitario en Economía DEClDE y colaboradora docente en el Departamento de Economía Aplicada y Estadística de la UNED. Anteriormente ha trabajado en Evaluación en ICEX España y como consultora junior. Sus intereses de investigación giran en torno a la identificación de causalidad entre comportamiento y preferencias políticas, y el contexto socio-económico. En concreto, ha trabajado sobre temas de polarización y movilidad social, y se encuentra investigando sobre normas sociales y discriminación partidista con métodos experimentales.

#### RESUMEN

Existe una preocupación creciente por la efectividad de las políticas públicas, tanto a nivel académico como práctico. El análisis de políticas públicas es cada vez más técnico y requiere de múltiples técnicas

---

<sup>1</sup> Los autores agradecen a los revisores anónimos y a los editores de la revista GAPP sus comentarios y sugerencias sobre la versión inicial del artículo. Sus contribuciones han sido clave para mejorar y completar la versión final.

para establecer de forma robusta el efecto causal de una política. En este artículo, repasamos algunas de las técnicas basadas en la experimentación más adecuadas para este fin. Examinaremos el uso de experimentos aleatorizados tanto de laboratorio como de campo, experimentos virtuales o experimentos naturales. A su vez, abordamos también algunos problemas habituales en el trabajo experimental en políticas públicas, como la validez de los resultados o la ética. A pesar de los posibles inconvenientes, la incorporación de la lógica experimental al análisis de políticas públicas es en la actualidad el patrón de referencia en las principales instituciones internacionales, y así empieza a ser también en España. El manejo de las técnicas que aquí repasamos, así como entender su importancia y limitaciones, se hace imprescindible para cualquier persona interesada en el ámbito de las políticas públicas. Este artículo sirve de introducción y guía en ese aprendizaje. Añadiendo la experimentación a nuestros análisis, podremos obtener resultados robustos sobre los que construir futuras políticas más eficientes y efectivas.

## **PALABRAS CLAVE**

Experimentos; evaluación de políticas públicas; causalidad; validez; ética de la experimentación.

## **ABSTRACT**

There is increasing concern about the effectiveness of public policies, both at the academic and practical level. Public policy analysis gets more technical every day, and multiple tools are required to establish the causal effect of a policy. In this article, we present an overview of some of the most suitable experimentation-based techniques for this purpose. We examine laboratory and field randomised experiments, on-line experiments and natural experiments. We also address some of the most common caveats of using experiments in public policies, such as validity of results or ethics. Despite possible hassles, the inclusion of experiments in public policy analysis has become the norm in international organisations, and that process is also starting in Spain. Mastering the techniques we present here, as well as understanding their relevance and limitations, is essential for anyone interested in the field of public policy. This article serves as an introduction and guide in that learning. Including experimentation in public policy analysis, we can obtain robust results upon which build more efficient and effective policies.

## **KEYWORDS**

Experiments; evaluation of public policy; causality; validity; experimentation ethics.

## **SUMARIO**

1. INTRODUCCIÓN. 2. ¿QUÉ CARACTERIZA A LA EVALUACIÓN EXPERIMENTAL DE POLÍTICAS PÚBLICAS? 2.1. ¿POR QUÉ EVALUAR EXPERIMENTALMENTE? 2.2. ¿QUÉ ES UN EXPERIMENTO? 2.3. ¿PARA QUÉ QUEREMOS HACER UNA EVALUACIÓN EXPERIMENTAL? 3. TIPOS DE EXPERIMENTOS EN POLÍTICAS PÚBLICAS. 4. LA VALIDEZ DE LOS EXPERIMENTOS EN POLÍTICAS PÚBLICAS. 5. LA ÉTICA EN LA EVALUACIÓN DE POLÍTICAS PÚBLICAS. 6. EL FUTURO DE LA EVALUACIÓN DE POLÍTICAS PÚBLICAS. 7. CONCLUSIÓN. REFERENCIAS BIBLIOGRÁFICAS.

## **1. INTRODUCCIÓN**

Los métodos experimentales han llegado a las ciencias sociales para quedarse. Lo que durante la segunda mitad del siglo xx fueron intentos relativamente localizados de aplicación de la experimentación en las diferentes disciplinas sociales, es hoy en día un cuerpo coherente de investigación sólida y replicable, y un acervo cada vez mayor de recursos metodológicos. La trayectoria en las diferentes disciplinas (economía, sociología, ciencia política) ha sido diferente, pero en todas ellas la experimentación no solo está plenamente integrada, sino que es central en algunas de las investigaciones más novedosas y de mayor repercusión (Gereke y Gërkhani, 2019; Gërkhani y Miller, 2022).

La proliferación del uso de los experimentos en ciencias sociales, así como la creciente sofisticación de las técnicas empleadas, ha complicado el aprendizaje y la enseñanza del análisis de políticas públi-

cas. Al mismo tiempo, esta expansión también se ha visto reflejada en el desarrollo de material docente y la oferta de cursos disponible, especialmente a nivel de posgrado. En cuanto a los manuales, Morton y Williams (2010) sigue siendo el texto de referencia para la ciencia política experimental. Se trata, además, de una fuente muy apropiada para obtener los fundamentos metodológicos para la evaluación de políticas públicas dado su especial énfasis en las técnicas de análisis causal y estadístico. En economía, la mayoría de cursos sobre la evaluación de impacto de las políticas públicas usan como libro de cabecera el manual de econometría de Angrist y Pischke (2009). En sociología, el libro reciente de Barrera et al. (2023) pone el énfasis en los diferentes diseños experimentales que pueden ser utilizados en la práctica, algo que resulta de utilidad para el análisis de políticas públicas de forma comprehensiva, no solo su impacto. En cuanto a la docencia, en la mayoría de facultades de economía ya se pueden encontrar cursos de posgrado sobre economía experimental y, en muchas, se ha incorporado la enseñanza de la experimentación en el grado. En el caso de la ciencia política tampoco es ya infrecuente encontrar cursos de posgrado en experimentos en ciencia política<sup>2</sup>.

En el espectacular desarrollo de la experimentación en ciencias sociales, la aplicación al análisis y evaluación de políticas públicas ha desempeñado un papel fundamental. En primer lugar, porque cuando la investigación experimental era solo marginal en las diferentes disciplinas sociales, ya se empleaban profusamente experimentos en el campo de las políticas públicas. De hecho, algunos de los ejemplos pioneros del uso de la experimentación social no se dieron en el interior de fríos laboratorios, sino al calor de intervenciones gubernamentales sobre aspectos como las políticas de empleo, los subsidios públicos o la vivienda (Oakley, 1998). En segundo lugar, la experimentación social ha adquirido el prestigio académico y público que hoy tiene precisamente por su aportación al análisis de políticas públicas, sobre todo en el caso concreto de los experimentos de campo aleatorizados. Este tipo de experimentos han sido adoptados como estándar de evaluación de políticas por instituciones multilaterales como el Banco Mundial<sup>3</sup>, y fueron reconocidos con el Premio Nobel en Economía en 2019 para los investigadores Esther Duflo Abhijit Banerjee y Michael Kremer<sup>4</sup>. Gracias a ellos la evaluación del impacto de las políticas públicas predominante es la experimental, y podemos encontrar cada vez más textos específicos sobre el uso de experimentos de en políticas públicas como Weiss y Birckmayer (2009), John (2017) o más recientemente el número especial editado por Akerlof, Oliver y Sunstein (2021).

En este artículo, repasamos algunos de los conceptos y fundamentos básicos sobre el uso de experimentos en políticas públicas, así como los principales tipos de experimentos, con el objetivo de ofrecer una guía en la enseñanza de estas técnicas. A lo largo del artículo desgranaremos qué debe entenderse por análisis experimental de políticas públicas, pero antes es importante diferenciar entre qué se entiende y se enseña como evaluación de políticas públicas en economía o en una facultad de ciencias políticas o administración. Esta distinción iluminará algunos de los conceptos que introduciremos más adelante en el texto. En los cursos de gobierno y administración pública, se aborda el análisis de políticas públicas desde una perspectiva amplia, que comprende todas las fases y pormenores del desarrollo de una política. Desde esta perspectiva, se busca analizar las dinámicas que actúan desde la propia concepción de la política y sus objetivos, hasta la implementación o posterior evaluación. Para ello, se incluyen tanto técnicas de análisis cuantitativo como cualitativo, y diseños *ex ante* (con anterioridad a la aplicación de la política) o *ex post* (con posterioridad a su aplicación)<sup>5</sup>. Por el contrario, en una escuela de economía, la evaluación de políticas públicas se centra casi exclusivamente en la enseñanza de técnicas estadísticas de análisis causal, comenzando por los experimentos aleatorizados<sup>6</sup>. Esta estrategia persigue aislar de manera robusta el efecto concreto que haya podido tener una política sobre

<sup>2</sup> Un ejemplo es el curso “Experimentos y comportamiento político” impartido en el Máster Universitario en Análisis Político y Electoral de la Universidad Carlos III de Madrid. El programa puede ser consultado aquí: <https://aplicaciones.uc3m.es/cpa/generaFicha?est=344&sig=17747&idioma=1>

<sup>3</sup> El manual *La Evaluación De Impacto en la Práctica* del Banco Mundial (Gertler et al., 2016) es un importante texto de referencia en la materia. Puede consultarse en <https://openknowledge.worldbank.org/handle/10986/25030>

<sup>4</sup> Muchas de las técnicas y aprendizajes desarrollados por estos investigadores y otros colaboradores pueden encontrarse en la guía de Glennerster y Takavarasha (2013).

<sup>5</sup> Un ejemplo puede encontrarse en el curso “Evaluación aplicada a políticas y programas públicos” del Máster en Gobierno y Administración Pública de la Universidad Complutense de Madrid. <https://www.ucm.es/estudios/master-gobiernoyadministracionpublica-plan-603619>

<sup>6</sup> Un ejemplo puede encontrarse en el curso “Empirical Methods for Public Policy Evaluation” del Centro de Estudios Monetarios y Financieros (CEMFI). [https://www.cemfi.es/programs/css/course\\_previous\\_years.asp?c=15&y=2021](https://www.cemfi.es/programs/css/course_previous_years.asp?c=15&y=2021)

indicadores objetivo. Nuestro repaso al uso de experimentos en políticas públicas se acercará más a la visión más cuantitativa y centrada en la causalidad que tienen los economistas, donde se han explorado más ampliamente las ventajas de la experimentación. Con todo, también repasaremos las posibles limitaciones de esta perspectiva.

## **2. ¿QUÉ CARACTERIZA A LA EVALUACIÓN EXPERIMENTAL DE POLÍTICAS PÚBLICAS?**

### **2.1. ¿Por qué evaluar experimentalmente?**

Antes de definir qué es un experimento y por qué es útil en materia de políticas públicas, repasaremos algunos conceptos necesarios en el proceso de evaluación de una política pública. El primer paso ha de ser siempre identificar y definir tres elementos clave. En primer lugar, debe delimitarse cuál es la política objeto de estudio, que también llamaremos tratamiento o intervención. Se trata generalmente de una actuación que podemos considerar externa, y el objetivo de la evaluación será determinar qué efectos o cambios ha podido tener sobre una población determinada. En segundo lugar, se debe definir la población objetivo, tanto en términos de límites geográficos como en características sociodemográficas. Finalmente, el tercer elemento fundamental a delimitar son las variables o dimensiones sobre las que queremos que la política actúe o creamos que lo va a hacer. Esta definición de variables objetivo es importante para el diseño de la política, que será diferente en función de los efectos que se pretendan lograr, pero también para determinar si ha tenido impacto o no. Puede darse el caso de políticas que no cumplan el objetivo para el que fueran creadas, pero que sí produzcan otros efectos positivos no anticipados.

Una vez delimitadas claramente la política a estudiar, la población objeto de estudio y las variables objetivo, podemos plantear el análisis. Hemos afirmado que nos aproximamos al estudio de las políticas públicas principalmente desde una perspectiva centrada en determinar el impacto causal de una determinada política. Desde este enfoque, el análisis o evaluación de políticas busca responder a la pregunta: ¿qué efecto ha tenido la implementación de esta política? Para contestar a esta pregunta, deberíamos observar cambios en las variables objetivo después de la implementación y asegurarnos que esos cambios se deben a la intervención y no a otros factores. En un mundo ideal, esto podría hacerse observando dos universos paralelos cuya única diferencia fuera la implementación de la política bajo estudio. Observando las diferencias entre esas dos realidades, sabríamos exactamente qué efectos ha tenido nuestra política. Sin embargo, al menos por el momento, es imposible observar varias realidades paralelas. Esta imposibilidad de conocer qué habría pasado sin la implementación de la política a estudiar se conoce como el problema del contrafactual y lleva décadas ocupando a los investigadores<sup>7</sup>. Las técnicas experimentales son hasta la fecha la herramienta más exitosa para solventar este problema, por su robustez y buenos resultados.

### **2.2. ¿Qué es un experimento?**

Un experimento en políticas públicas es una intervención con unas características concretas que permite informar de forma correcta sobre los efectos de dicha intervención. Pensemos a modo de ejemplo en una política dirigida a mejorar la inclusión digital de un colectivo vulnerable. Para empezar, delimitamos los tres elementos clave que definen esa intervención: la población objetivo de la intervención serán personas mayores de cincuenta años con bajo nivel educativo, la intervención tomará la forma de un subsidio para la compra de material informático y la inclusión digital será medida como el número de interacciones digitales con la administración y la calidad de las mismas. Para evaluar el efecto de esta política se podría recurrir a un análisis observacional tradicional, midiendo el número de interacciones digitales de estas personas antes y después de la intervención o comparando las interacciones de quienes han recibido el subsidio con aquellos que no. Sin embargo, esta estrategia observacional no es

---

<sup>7</sup> Dos estrategias habituales para aproximarse a este problema y determinar el efecto de una intervención son comparar dos momentos diferentes en el tiempo (antes vs. después de la intervención) o a dos grupos diferentes de personas (grupo tratado vs no tratado). Estas estrategias, aunque de sencilla implementación, presentan una serie de problemas técnicos que dan lugar a sesgos y resultados poco fiables. Se puede encontrar una discusión más amplia sobre falsas estimaciones del contrafactual en Gertler et al. (2016).

suficiente para aislar los efectos de la intervención y asegurar la causalidad. Podría darse el caso que solo las personas con más habilidades digitales soliciten la ayuda, o que haya habido otras políticas implementadas en ese tiempo que tengan efectos que se mezclen con la intervención a estudiar. Para hacer de esta intervención un experimento, son necesarias dos características definitorias: la manipulación y el control.

En general, las teorías preexistentes o las experiencias previas permiten a los evaluadores (también a los responsables de formular políticas) tener una serie de hipótesis o expectativas sobre cómo funciona cierta política y cuáles pueden ser los resultados de la evaluación, es decir, acerca de las relaciones causales entre la variable sobre la que se interviene, o tratamiento, y las variables de resultado, o variables objetivo. En nuestro ejemplo de la política para la inclusión digital, esperaríamos que un subsidio para la compra de material informático facilite el acceso a nuevos dispositivos y, por tanto, aumente las interacciones digitales. El objetivo de un trabajo de evaluación es comprobar hasta qué punto estas expectativas se cumplen empíricamente. Para ello, la ventaja de la metodología experimental es que nos permite alterar discrecionalmente características de la variable explicativa, esto es, del tratamiento o intervención que se considera la causa, y observar a continuación el efecto de este cambio en la variable de resultado, la variable dependiente u objetivo. Estos cambios intencionales sobre la intervención son lo que llamamos manipulación experimental. Por otra parte, el investigador debe garantizar que cualquier resultado observado se deba a la manipulación experimental y no a otros factores. Es aquí donde entrará el control en juego, ya que un experimento habrá sido diseñado siempre de manera que cualquier otro efecto que pueda alterar el resultado quede bajo control, aislando del contexto y factores externos.

Siguiendo el ejemplo propuesto, para hacer de la intervención un experimento sería necesario realizar algún tipo de manipulación sobre cómo se asigna o reparte el subsidio para comparar la inclusión digital de aquellas personas que han recibido el tratamiento con las que no. Solo así se podría observar el impacto de la intervención. Sin embargo, como puede haber otros factores que afecten a la inclusión digital además de la intervención, garantizaremos la comparabilidad entre los grupos tratado y no tratado controlando las personas que conforman cada grupo, asegurando homogeneidad entre ellos. Es importante notar que las condiciones que van a permitir la evaluación posterior deben diseñarse con anterioridad a la puesta en marcha de la política. En otras palabras, el investigador o evaluador *interviene* en la realidad *manipulando* la implementación y ejerciendo un *control* sobre la generación de los datos. En los estudios observacionales no existe dicho control y manipulación.

Como decíamos más arriba, determinar el efecto causal de la variable independiente sobre el resultado depende de la capacidad que tengamos para afirmar que ese cambio se ha producido exclusivamente como fruto de nuestra manipulación y no hay otros factores interviniendo en la situación. En ciencias como la física esto se garantiza mediante el control absoluto de todos los parámetros que puedan afectar al experimento, como la presión, la gravedad o la velocidad. Sin embargo, un control a este nivel no es posible en las ciencias sociales, donde trabajamos con personas que cuentan con sus propias actitudes, percepciones, rutinas y experiencias. Por tanto, el control experimental se ejerce mediante la asignación de los participantes. No controlar correctamente esta asignación puede suponer, entre otros problemas, la inclusión de sesgos indeseados. ¿Cómo podemos solucionar entonces el problema del sesgo introducido por la selección de participantes? La solución más habitual es la aleatorización.

Muchos manuales comienzan con la aleatorización como el elemento central de la evaluación experimental de políticas públicas. De hecho, los experimentos sociales más usados en evaluación suelen ser también denominados experimentos controlados aleatorizados (o randomized controlled trials, abreviado RCT, en inglés). Nosotros hemos preferido comenzar por el verdadero objetivo de los diseños experimentales, el control directo, siendo la aleatorización solo un sustituto en tanto que nos proporciona un control indirecto (Rubin, 1974). Mientras que un control directo nos permitiría conocer y manipular cada una de las variables que intervienen en una evaluación, pudiendo observar cualquier característica introducida por los participantes, lo que nos ofrece la asignación aleatoria de estos a las condiciones experimentales es que las características de los participantes en cada tratamiento son, en media, similares. Si nuestro subsidio para la inclusión digital fuera solicitado principalmente por personas de entornos urbanos o con hijos jóvenes que les puedan ayudar, y observamos después que estas personas efectivamente tienen más interacciones digitales con la administración que otras de su edad y nivel educativo, no podremos afirmar empíricamente que la diferencia se deba a la intervención o, alternativamente, a la ayuda de los hijos. Sin un control sobre la homogeneidad entre los que reciben el subsidio y aquellos con los que están siendo comparados, no podremos afirmar que la intervención es efectiva. En cambio, si tanto el

grupo tratado como el no tratado son homogéneos y similares en las características que los definen, se diluye el efecto de posibles sesgos. Esta homogeneidad se consigue asignando aleatoriamente las personas que reciben el tratamiento entre aquellas que cumplan los requisitos. La imposibilidad de controlar factores muchas veces no observables, como la habilidad o la experiencia personal, es lo que ha hecho que los científicos sociales pongan el peso del diseño de las evaluaciones en la asignación aleatoria de participantes a los tratamientos.

El foco casi exclusivo en la aleatorización, que no deja de ser una herramienta, hace que en la enseñanza de los métodos experimentales se pase por alto que el objetivo último es el control y la manipulación. Pero, además, produce otro efecto, y es que la asignación aleatoria no resuelve todos nuestros problemas a la hora de diseñar un experimento adecuado. Un problema que puede pasar desapercibido es la motivación de los participantes experimentales una vez que los hemos asignado a los tratamientos y el control. Esto es especialmente relevante en muchas evaluaciones de políticas en las que el grupo de control se define como aquel que no recibe la intervención. En una evaluación experimental, es importante tener un grupo de control sobre el que comparar los resultados. Sin embargo, puede ocurrir que el grupo que recibe la política pública, sabiéndose observado o atendido, ve alteradas sus motivaciones y comportamientos, de manera que los resultados no sean realmente comparables. Puede que un subsidio por sí solo no sea suficiente para mejorar la inclusión digital de colectivos vulnerables, pero los individuos seleccionados, sabiendo que van a ser evaluados, hagan un esfuerzo adicional que haga parecer que la política es efectiva, y que no se repetiría si se implementa a gran escala. En un ensayo clínico, por ejemplo, resuelven este problema usando un placebo, de manera que la experiencia durante el experimento es idéntica para el grupo tratado y el control y las diferencias solo pueden deberse al medicamento testado. Sin embargo, es mucho más difícil en el ámbito de las políticas públicas imaginar una «intervención placebo». En estos casos, las soluciones adoptadas por los investigadores son variadas, desde la compensación económica de todos los participantes, como ocurre en muchos experimentos económicos, hasta la posibilidad de que el grupo de control reciba la intervención en un momento posterior en el tiempo, una vez que la evaluación ya haya concluido.

### 2.3. ¿Para qué queremos hacer una evaluación experimental?

Muchos experimentos, en el ámbito de las ciencias sociales y de la ciencia en general, se hacen para responder a un *porqué*: ¿por qué los ciudadanos pagan impuestos?, ¿por qué las minorías son discriminadas en el mercado de trabajo?, ¿por qué donamos dinero a fines sociales de forma voluntaria? Todas estas preguntas requieren de cierto conocimiento previo del problema en cuestión, el cual nos permite formular hipótesis como las que referimos en el apartado previo. Es lo que en la literatura se conoce como experimentos de orientación teórica (Miller, 2006). Sin embargo, en el ámbito del análisis de políticas públicas este no suele ser el caso y, en realidad, estamos en muchas ocasiones interesados en preguntas puramente descriptivas, como: ¿afecta la cuantía del subsidio de desempleo a la incorporación de personas desempleadas al mercado de trabajo?, ¿aumentan las clases de refuerzo el rendimiento de los estudiantes de secundaria? En este segundo tipo de preguntas no siempre tenemos evidencia previa suficiente que nos permita formular hipótesis precisas. A este segundo tipo de experimentos se les conoce como experimentos de orientación empírica (Barrera et al., 2023) e investigan regularidades empíricas para las que no existe una explicación clara por parte de las teorías existentes.

Que un experimento tenga finalidad teórica o empírica tiene consecuencias importantes sobre el modo en el que se deben interpretar sus resultados. Por ejemplo, si un investigador diseña un experimento para comprobar si las tutorías presenciales tienen un mayor impacto sobre el rendimiento de los alumnos que las virtuales, debe argumentar por qué cree que este es el caso, incluso si lo que le interesa es solamente la relación empírica entre tutorías y rendimiento. Si no se hace explícito el mecanismo que subyace a esta relación difícilmente aprenderemos algo que vaya más allá del caso concreto que nos ocupa. Los experimentos de orientación teórica nos informan sobre los mecanismos subyacentes en una intervención y, si bien no tienen como objetivo determinar el impacto de una política, nos permiten ahondar en por qué una política puede tener o no efecto y mejorar el diseño de futuras políticas y evaluaciones. Es importante tener en cuenta que establecer el efecto causal de una intervención sobre una determinada variable que queremos medir no es una cuestión técnica sino de credibilidad. Cuanta más evidencia tengamos acerca de un determinado fenómeno, más creíble será el resultado que queremos mostrar.

### 3. TIPOS DE EXPERIMENTOS EN POLÍTICAS PÚBLICAS

La mayoría de evaluaciones experimentales de políticas públicas toman la forma de un experimento de campo, es decir, de una intervención directa sobre el contexto y las personas objeto de la evaluación. Sin embargo, este no es el único tipo de diseño experimental posible y, especialmente en el contexto de la docencia de los métodos experimentales, es importante repasar otras tipologías de diseño. Los tipos de experimentos más comunes son los de laboratorio, de campo, encuestas factoriales y experimentos naturales. En este apartado, presentaremos cada uno de ellos poniendo el énfasis en sus fortalezas y debilidades.

Los *experimentos de laboratorio* son probablemente el tipo de experimento más conocido, incluso en las ciencias sociales (Falk y Heckman, 2009; Webster y Sell, 2014). Contienen los tres elementos comunes al resto de experimentos y que hemos introducido con anterioridad –manipulación, comparación grupal y aleatorización–. Tienen lugar en el contexto relativamente artificial, anónimo y controlado del laboratorio. Llamamos laboratorio a un entorno de investigación diseñado a propósito para realizar experimentos, aislado de posibles perturbaciones externas que puedan alterar o enmascarar el efecto que se pretende estudiar. La principal fortaleza del laboratorio es que permite al investigador tener un control preciso y estricto sobre el entorno, la manipulación experimental y la medida de las variables independientes y dependientes. Debido a este alto grado de control, y a la posibilidad de replicar los resultados obtenidos en otros contextos o por otros investigadores, el laboratorio es la localización prototípica para la realización de experimentos. Sin embargo, muchos investigadores en el campo de las políticas públicas son escépticos ante la posibilidad de generalizar los resultados generados en el laboratorio a situaciones reales de interés. Esta crítica se refiere sobre todo al carácter artificial del laboratorio y a la práctica habitual de utilizar muestras de conveniencia<sup>8</sup>, como estudiantes universitarios. A pesar de ello, los experimentos de laboratorio son especialmente útiles para comprobar hipótesis teóricas a partir de las que diseñar políticas públicas más efectivas o para probar versiones piloto de una intervención antes de llevarlas a un contexto menos controlado. Valga como ejemplo la investigación de laboratorio sobre el comportamiento de hombres y mujeres en entornos competitivos (Schram et al., 2019; Gërzhani, 2020; Gërzhani et al., 2021). Esta muestra que, a la hora de resolver tareas en un contexto de competición, las mujeres tienden a tener peores resultados que los hombres, mientras que en contextos de no competición los resultados son equivalentes. Si bien estos experimentos no evalúan una política concreta, analizan mecanismos de comportamiento omnipresentes que pueden afectar al desempeño en cualquier intervención que suponga una competición por estatus o recursos. Si planteamos una formación para personas desempleadas, quizá pensemos que añadimos una motivación extra si esta formación incluye una evaluación o incluso un premio final para los mejores, pero puede que estemos consiguiendo efectos diferenciados por género sin preverlo. Este tipo de heterogeneidades solo pueden ser observadas y explicadas en el contexto controlado del laboratorio, y es por ello que los experimentos de laboratorio son un complemento muy valioso al resto de experimentos que se pueden usar en el ámbito de las políticas públicas.

Un segundo grupo importante de experimentos, probablemente el más usado en la evaluación de políticas, es el de los *experimentos de campo* (Baldassarri y Abascal, 2017; Gerber y Green, 2012; Harrison y List, 2004; Al-Ubaydli et al., 2021). Los experimentos de campo son llevados a cabo en un entorno «natural» donde los participantes en los mismos interactúan en situaciones normales de su vida cotidiana al realizar la tarea objeto de investigación. El hecho de que se realicen en este contexto natural es lo que los distingue de otros tipos de experimentos, y es la razón principal por la que muchos investigadores los emplean para el análisis y la evaluación de políticas públicas. Permiten incorporar el contexto social, investigando a personas en condiciones «naturales» y recogiendo las medidas del experimento sin que los participantes sean conscientes de ello. Al participar en un experimento de campo oculto (concealed), los participantes no son conscientes de que participan en la investigación, lo que reduce sustancialmente una de las mayores amenazas de la investigación experimental: la reacción de los participantes tras conocer que son objetos de investigación (Shadish et al., 2002). En contra del entorno controlado del laboratorio, en el campo muchos factores pueden influir sobre el desarrollo de la investigación, sin posibilidad

<sup>8</sup> Las denominadas muestras de conveniencia responden a un procedimiento de selección de participantes en la investigación que prima la disponibilidad de la muestra y no la representatividad, la diversidad u otra característica deseable de la misma. El ejemplo más común de muestra de conveniencia son los estudiantes universitarios que dan su consentimiento para ser registrados y acudir, a veces de forma repetida, a estudios experimentales.

de ser contenidos y a menudo difíciles de medir. Entre estos problemas debemos incluir las desviaciones de la asignación experimental, el contagio entre participantes en los tratamientos y en el control, el rechazo a tomar parte en un tratamiento, y la «muerte experimental», es decir, el abandono de participantes antes de que concluya el experimento. Todos estos problemas implican que la línea que separa a personas tratadas del grupo de control se diluye y se presentan inexactitudes. En el mejor de los casos, en el que estas desviaciones se dan de forma desordenada e independiente de cualquier factor relevante, el efecto de la política será infraestimado. Sin embargo, puede que las personas que abandonen el estudio o que no sigan su asignación presenten alguna característica en común que les diferencia del resto, o simplemente la política no está siendo efectiva para ellos. En estos casos, el resultado de la evaluación será erróneo. Por este motivo, al analizar un experimento de campo es especialmente importante no solo observar que la asignación al tratamiento se hace con criterios experimentales, habitualmente mediante aleatorización, sino también que dicha asignación se mantiene a lo largo de toda la intervención o, al menos, que las desviaciones sean pocas y no introduzcan sesgos. Además, pueden existir restricciones éticas, legales, políticas y prácticas (por ejemplo, la mera posibilidad de realizar un experimento en condiciones naturales) que supongan retos importantes para la realización del experimento tal y como ha sido diseñado. A pesar de estas limitaciones, la posibilidad de realizar un experimento controlado y aleatorizado en el entorno real en el que se hallan cotidianamente los participantes hacen a los experimentos de campo la técnica de investigación experimental predominante en la evaluación de políticas públicas, ya que permite analizar una intervención en su entorno de implementación real. Algunos investigadores han denominado a estos experimentos el *gold standard* de la investigación experimental (Al-Ubaydli y List, 2015). Una evaluación de este tipo solo se podrá llevar a cabo si se ha implementado la política de acuerdo a los criterios experimentales necesarios y, por tanto, deberá tenerse en cuenta desde la fase inicial de diseño de la intervención.

Uno de los ejemplos más conocidos y pioneros en el uso de metodología experimental es el del programa contra la pobreza PROGRESA (Gertler, 2004; Parker y Teruel, 2005). En 1997, el gobierno mexicano implementó un programa de transferencias condicionadas con el que perseguía mejorar las condiciones de las personas en situaciones de pobreza y reducir su transmisión intergeneracional. El despliegue de PROGRESA se hizo eligiendo localidades con altos niveles de pobreza, y se asignó aleatoriamente cuáles de estas localidades serían las primeras en recibir el programa. De esta manera, una restricción presupuestaria se puede convertir en la manipulación necesaria para evaluar experimentalmente una intervención. En este caso, asignar el programa a familias escogidas aleatoriamente no sería una buena opción, ya que existe un alto riesgo de que ciertos efectos sobre, por ejemplo, la salud o la escolarización de los menores se transmitan entre familias de una misma comunidad, reciban o no la transferencia. La aleatorización a nivel de localidad evita estos problemas. En España, la implementación de políticas que incluyan una manipulación experimental para su evaluación es aún muy limitada, pero ya empezamos a encontrar algunos ejemplos localizados.

Un tipo específico de experimentos de campo serían los *experimentos virtuales* (online). La mayoría de personas en los países industrializados tienen acceso a internet, usan teléfonos inteligentes y son activos en una o varias redes sociales. Estas tres tendencias proporcionan nuevas oportunidades para llevar a cabo experimentos, incluidos aquellos cuya finalidad es el análisis y la evaluación de políticas públicas (Salganik, 2019). Los experimentos virtuales permiten obtener datos a gran escala de poblaciones mucho más diversas que las muestras de conveniencia utilizadas en los experimentos de laboratorio y algunos experimentos de campo. Al mismo tiempo, posibilitan un nuevo tipo de control relacionado con el seguimiento de las actividades virtuales de los participantes en tiempo real. Debido a su bajo coste y al enorme tamaño potencial de la muestra, estos experimentos son ideales para poner a prueba diseños complejos que combinan un número elevado de tratamientos, así como para explorar la interacción entre estos tratamientos y las características personales de los participantes.

Los experimentos virtuales comparten algunas de las fortalezas y debilidades de los experimentos de campo más tradicionales. Son llevados a cabo en un entorno «natural» y, por tanto, permiten estudiar el comportamiento social en un entorno menos abstracto y artificial que en los de laboratorio, pero sufren también de una potencial falta de control que puede afectar a la identificación del efecto que se pretende estimar. Por ejemplo, en muchos experimentos virtuales no es posible saber a ciencia cierta qué persona o grupo de personas son las que están participando realmente en el experimento. El terreno de los experimentos virtuales es relativamente nuevo y aún estamos lejos de conocer su verdadero potencial, así como los retos metodológicos que plantea.



Un tercer tipo de experimentos son los *experimentos multifactoriales en encuestas*, que incluyen los estudios con viñetas, los experimentos de decisión (choice experiments) y los análisis conjuntos (conjoint analysis) (Mutz, 2011). En los experimentos multifactoriales en encuestas, los participantes se enfrentan a situaciones hipotéticas de decisión y se les pide que evalúen diferentes escenarios o que elijan entre distintas alternativas. Cada escenario o alternativa varía sistemáticamente entre tratamientos y son asignados aleatoriamente a los participantes. En este sentido, combinan el rigor metodológico de un diseño experimental con las ventajas de la investigación por encuesta. Al contrario que en los experimentos de laboratorio, en los experimentos multifactoriales en encuesta es más fácil y barato obtener datos a gran escala necesarios para desarrollar diseños factoriales complejos con muchos tratamientos. Al incluir un módulo experimental en una encuesta, los investigadores también tienen acceso a realizar estudios con muestras representativas que incrementan la generalización de los resultados. Dado que la manipulación experimental ocurre solo para escenarios hipotéticos, no sufren las limitaciones éticas y prácticas de los experimentos de laboratorio y de campo. Por tanto, esta aproximación ofrece un método flexible para la evaluación hipotética de diferentes intervenciones de políticas, por ejemplo, investigando cuestiones relacionadas con la aceptabilidad de políticas públicas, sin el coste ni los riesgos de la puesta en práctica real de la política en concreto. Pero, como en el resto de tipos, los experimentos en encuestas también tienen algunas desventajas. Cuestiones típicas de las encuestas, como la deseabilidad social de algunas respuestas o la reacción que preguntas y respuestas producen en los encuestados, pueden suponer un problema cuando los tratamientos experimentales son demasiado obvios. Además, siempre queda la duda de hasta qué punto las intenciones reveladas ante escenarios hipotéticos se corresponden con el comportamiento en el mundo real. Con todo, los experimentos multifactoriales en encuesta son especialmente útiles en la fase de diseño de políticas para informar sobre las preferencias de los ciudadanos de una manera completa. Caparrós et al. (2007) usan un experimento de decisión para estudiar el valor otorgado a la reforestación del Parque Natural Los Alcornocales con alcornocales o con eucaliptos, integrando en la decisión cuestiones sobre la biodiversidad, el valor de mercado o la absorción de CO<sub>2</sub>. Este tipo de experimentos son clave para escoger entre intervenciones alternativas que cubren una misma necesidad, ya que los individuos no solo expresan sus preferencias respecto al resultado, sino que integran el análisis de los costes de cada opción a la hora de elegir. En un mundo en el que los recursos disponibles son limitados, elegir qué intervenciones queremos implementar es tan importante como asegurarse de que estas sean efectivas.

Por último, nos detendremos en otro de los diseños clásicos en la evaluación de políticas públicas: los *experimentos naturales*. En realidad, este tipo de diseño no trabaja con datos experimentales en el sentido descrito más arriba, ya que, en vez de manipular activamente un tratamiento, los investigadores aprovechan variaciones ocurridas en contextos naturales para intentar estimar algún efecto causal (Dunning, 2012). Estas variaciones exógenas al diseño de investigación a veces se producen fruto de eventos naturales (como los efectos de un huracán) o de una decisión humana (como un ataque terrorista). Los experimentos naturales no son experimentos en un sentido estricto, el experimentador no tiene control sobre la manipulación experimental ni sobre la asignación de participantes, pero los fundamentos metodológicos de los experimentos naturales con su asignación cuasi-aleatoria los hacen un método comparable al resto de experimentos. Este supuesto de asignación cuasi-aleatoria es el punto más crítico de este tipo de diseños, y será responsabilidad de los investigadores comprobarlo y darle credibilidad. Aun así, estos diseños se han convertido en una herramienta muy utilizada en la evaluación de políticas, ya que los cambios de políticas o las heterogeneidades en su implementación son normalmente utilizados como variación exógena que define el tratamiento.

Algunos ejemplos pueden ayudarnos a mostrar la contribución de los experimentos naturales. Nakamura et al. (2022) estudian el efecto que supone emigrar sobre el salario y la carrera profesional de la primera generación usando la erupción de un volcán como variación exógena. En 1973, un volcán entró en erupción en las islas Westman, Islandia, y la lava destruyó aproximadamente un tercio de los hogares, obligando a estas familias a mudarse fuera de la isla. El reto de los investigadores está en mostrar que las familias que emigraron tienen características similares a aquellas que no. Si, por el contrario, la lava hubiera destruido viviendas solo en una zona de ingresos altos de la isla, los resultados décadas después no sería comparables. Dinas et al. (2019), por su parte, usan la variabilidad en la exposición a la crisis de refugiados de diferentes islas griegas para estudiar el efecto de dicha exposición sobre el voto a partidos de extrema derecha, que promulgan mensajes anti-inmigración. La Tabla 1 proporciona un resumen de las principales ventajas e inconvenientes de los diferentes tipos de experimentos.

TABLA 1. PRINCIPALES TIPOS DE EXPERIMENTOS, VENTAJAS E INCONVENIENTES

	Ventajas	Inconvenientes
<b>Experimentos de laboratorio</b>	<ul style="list-style-type: none"> <li>Contexto muy controlado.</li> <li>Fácil replicabilidad.</li> </ul>	<ul style="list-style-type: none"> <li>Contexto artificial, normalmente poco realista (difícil extrapolar los resultados, baja validez externa).</li> <li>Normalmente, uso de muestras de conveniencia (estudiantes).</li> </ul>
<b>Experimentos de campo</b>	<ul style="list-style-type: none"> <li>Contexto natural, resultados muy robustos (con un buen diseño).</li> </ul>	<ul style="list-style-type: none"> <li>Difícil (o imposible) ejercer un control real o medir todos los factores externos que pueden interferir. Riesgo de «contagio» entre grupos, o abandono de los participantes. Limitaciones prácticas, éticas, legales o políticas.</li> </ul>
<b>Experimentos online</b>	<ul style="list-style-type: none"> <li>Contexto natural y fácilmente accesible para un amplio sector de la población gracias a los avances tecnológicos.</li> <li>Bajo coste.</li> </ul>	<ul style="list-style-type: none"> <li>Potencial falta de control, menor cuanto más natural sea el contexto.</li> <li>Discutible la extrapolación de los resultados fuera del entorno virtual.</li> </ul>
<b>Experimentos multifactoriales en encuestas</b>	<ul style="list-style-type: none"> <li>Fácil implementación mediante encuesta.</li> <li>Acceso a muestras representativas.</li> <li>Alta replicabilidad.</li> </ul>	<ul style="list-style-type: none"> <li>Presentación de escenarios hipotéticos, el reto es conseguir respuestas que se acerquen a los comportamientos en un escenario real.</li> <li>Potencial falta de control.</li> </ul>
<b>Experimentos naturales</b>	<ul style="list-style-type: none"> <li>Uso de datos observacionales, aprovechamiento de un suceso natural o un cambio producido de forma exógena.</li> </ul>	<ul style="list-style-type: none"> <li>No existe control. Se tendrá que argumentar cierta cuasi-aleatoriedad o conseguir técnicamente, aunque difícilmente demostrable.</li> </ul>

Fuente: elaboración propia.

#### 4. LA VALIDEZ DE LOS EXPERIMENTOS EN POLÍTICAS PÚBLICAS

En los apartados anteriores hemos mencionado otro de los temas que deben ser cubierto en la docencia de los métodos experimentales en políticas pública: la *validez* de la investigación. El problema de la validez en la investigación empírica tiene que ver con hasta qué punto podemos dar credibilidad a aquello que nos dicen los datos. En ciencias sociales, tradicionalmente se han utilizado las etiquetas de «validez interna» y «validez externa» de la investigación (Campbell y Stanley, 1963; Cook y Campbell, 1979). La *validez interna* se refiere a las inferencias que podemos hacer en el contexto de un único experimento y se pregunta hasta qué punto podemos afirmar que el efecto obtenido es un efecto *causal*. La *validez externa* ya no tiene que ver con la causalidad, sino con la posible generalización de los mismos a otros contextos distintos al del propio experimento donde hemos obtenido los resultados.

Aunque sea la forma más popular de contar los problemas de validez en la evaluación de políticas, la distinción entre validez interna y externa ha sido cuestionada ampliamente durante las últimas dos décadas (Jiménez-Buedo y Miller, 2010). Si uno se asoma a los libros de metodología experimental en psicología y ciencias sociales encontrará la idea de que existe una tensión entre la validez interna y externa de la investigación social. Para incrementar la validez interna, uno debe intentar aislar todos los factores que pueden afectar a la identificación del efecto que estamos poniendo a prueba, pero, al controlar todos estos factores y aislar el supuesto efecto, estamos, al mismo tiempo, creando una situación tan artificial que difícilmente será generalizable a otros contextos. Por el contrario, en entornos naturales, como generalmente se da en la evaluación de políticas públicas, uno quiere garantizar a toda costa el realismo de la intervención, pero, de este modo, se pierde el control experimental, tal y como planteábamos en el apartado anterior. Por tanto, nos encontraríamos ante el dilema de tener que elegir entre incrementar la validez interna o la externa de la investigación. Sin embargo, esta postura está cada vez más cuestionada, entre otros motivos porque la validez interna tendría preeminencia sobre la externa. Si no somos capaces de identificar el efecto causal de la intervención, para qué vamos a preocuparnos de su generalización. En realidad, parecería más bien que un incremento de la validez interna también contribuye a incrementar la externa. Independientemente de la toma de postura metodológica, resulta fundamental plantear esta cuestión en la docencia sobre métodos experimentales.

Un segundo tema central en la discusión de los problemas de validez en clase es cómo entender la generalización de los resultados experimentales. Aquí hay dos aproximaciones que han sido denominadas «aproximación empírica» y «aproximación teórica» a la validez (Barrera et al., 2023; Gërkhani y Miller, 2022). Esta distinción es muy importante para la evaluación de políticas públicas, ya que parte de un debate que se produjo precisamente a propósito de la aplicación de métodos experimentales a la evaluación de intervenciones, públicas y privadas. La «aproximación empírica» asume que los experimentos (de campo) controlados aleatorizados (RCTs) suponen el *gold standard* de la investigación experimental porque son tan buenos como los experimentos de laboratorio para estimar efectos causales (validez interna) y son casi siempre superiores a la hora de generalizar los resultados (validez externa) debido a que se producen en un contexto real. Esta aproximación empírica es la más común entre investigadores en ciencias sociales, incluidos aquellos que se dedican a la evaluación de políticas públicas. Lo que se intenta conseguir es que la investigación sea «realista», «similar» o «representativa» de la realidad que se pretende comprender. Por eso la aproximación de hacer experimentos aleatorizados en el mismo contexto social donde la realidad está ocurriendo parece una buena idea. En definitiva, la intervención solo consiste en asignar los participantes en intervenciones reales de un modo diferente: de modo aleatorio. Los ejemplos clásicos de intervenciones sociales aleatorizadas del siglo xx (Oakley, 1998) y las intervenciones del Banco Mundial, JPAL o IPA en la actualidad siguen esta estrategia<sup>9</sup>.

Sin embargo, en los últimos años se han levantado una serie de críticas frente a la supuesta validez superior de los RCTs sobre otros métodos observacionales y experimentales de investigación social. En un citado artículo, el también premio Nobel de economía Angus Deaton y la filósofa de la ciencia Nancy Cartwright (Deaton y Cartwright, 2018) cuestionan que los efectos causales obtenidos en un experimento concreto (por ejemplo, una intervención sobre familias desfavorecidas en Colombia) puedan ser directamente trasladados a otro contexto similar (familias desfavorecidas en India). En este sentido, que el diseño de un experimento garantice la validez interna, no nos garantiza que también alcance la validez externa. ¿Cómo podemos generalizar los resultados de una investigación según esta perspectiva? Pues a través de una teoría. Lo que dice esta «aproximación teórica» a la evaluación de políticas es que lo que nos enseña un experimento concreto es hasta qué punto la teoría que manejábamos con anterioridad tiene sentido en un contexto concreto (Colombia). Después, será la teoría, reformulada tras los resultados obtenidos en el primer experimento (Colombia), la que nos permita elaborar predicciones e hipótesis para nuestro nuevo experimento (en la India).

En última instancia las aproximaciones empírica y teórica al problema de la validez de los experimentos son complementarias y por eso ambas merecen un hueco en la docencia de los métodos experimentales. La aproximación empírica nos permite poner el foco en aspectos cuantitativos de la evaluación (cuántas familias mandan sus hijos a la escuela), mientras que la teórica lo hace más sobre aspectos cualitativos (cómo contribuye una ayuda económica a las familias a la escolarización de los hijos).

## 5. LA ÉTICA EN LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

El uso creciente de métodos experimentales ha puesto la ética de la investigación en el centro de la reflexión sobre los métodos de investigación en ciencias sociales, en general, y aún con más ahínco cuando se utiliza la experimentación en la evaluación de políticas públicas. ¿Por qué ha cobrado tal relevancia la ética de la investigación social en los últimos años? Desde hace décadas en los Estados Unidos, y de forma más reciente en España, las ciencias sociales han desarrollado marcos regulatorios e instituciones como los comités de ética a los que las instituciones científicas y las agencias financiadoras se ven sometidas. En los años 70 se establecieron en los Estados Unidos los *Institutional Review Boards* (IRB) y en 1979 el denominado Informe Belmont introdujo las líneas maestras de lo que desde entonces se entiende por ética de la investigación: consentimiento informado, evaluación de los riesgos y beneficios, y selección de los sujetos participantes en la investigación. En concreto: 1) los sujetos experimentales deben tener toda la información relevante acerca de su participación en la investigación de forma comprensible y su participación debe ser voluntaria; 2) los riesgos y beneficios de su participación deben ser evaluados de forma sistemática; 3) la selección de los participantes en la investigación debe ser el resultado de un proceso que distribuye los ries-

<sup>9</sup> Para más información ver: <https://www.povertyactionlab.org/es> y <https://www.poverty-action.org/>

gos y los beneficios de forma justa entre los participantes. En las ciencias sociales, todas las asociaciones profesionales de mayor relevancia (asociaciones americanas de sociología, psicología o economía) cuentan en la actualidad con sus propias recomendaciones sobre ética de la investigación. Una revisión exhaustiva de estas recomendaciones escapa con creces al objetivo de este artículo, por lo que nos centraremos en un tema concreto, dada su relevancia para la docencia y práctica de la evaluación experimental de políticas públicas: la prohibición del uso del «engaño» (deception) en las intervenciones experimentales.

Cuando hablamos de engaño nos referimos a proporcionar información falsa o incompleta a los participantes en un experimento (Hey, 1998). Para los estudiantes e investigadores que se acerquen por primera vez a los métodos experimentales, parecería que el uso del engaño, sea de forma directa o indirectamente a través de la ocultación de información, es algo que debería estar prohibido en la investigación social, especialmente en aquella auspiciada y financiada desde las administraciones públicas. Sin embargo, esta técnica ha sido ampliamente utilizada por psicólogos sociales y sociólogos durante décadas. Pensemos en los experimentos clásicos sobre conformismo social, donde los investigadores utilizan «compinches» (confederates) para inducir a los participantes en la investigación a pensar de un determinado modo. Hoy en día el uso del engaño no está aceptado salvo en investigaciones que no pudieran realizarse de otra forma.

De todos modos, aunque desde un punto de vista ético parecería que el engaño no tiene ninguna cabida en la evaluación experimental de políticas públicas, existe un área gris que es difícil de juzgar, tanto desde un punto de vista ético como práctico. Una información completa acerca del experimento podría suponer explicar a los participantes en la investigación incluso las hipótesis de la misma, algo que podría comprometer los propios resultados de la investigación. Siempre hay algún tipo de información que el investigador no proporciona, tratando de sostener un equilibrio complicado entre mantener la máxima transparencia con los participantes y no revelar información que comprometa los resultados. En la evaluación de políticas públicas este equilibrio es especialmente complicado, dado que generalmente esta se produce en un contexto real donde el control de toda la información que llega a los participantes es prácticamente imposible.

Aunque muchos programas avanzados de enseñanza de métodos experimentales incluyen sesiones sobre la ética de la investigación, ésta todavía ocupa un papel secundario frente a otras cuestiones, como el diseño de investigación o las técnicas de análisis. Dada la importancia de la ética para la protección de los participantes, la credibilidad del estudio y la reputación de los propios investigadores, la ética de la investigación debe ocupar un lugar destacado en la enseñanza de la evaluación experimental.

## 6. EL FUTURO DE LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

Hace ya unos años que la evaluación de políticas públicas está siempre presente en cualquier discusión sobre qué políticas queremos para el futuro. A lo largo del artículo hemos examinado las principales técnicas y sus problemáticas asociadas que definen la evaluación experimental como la conocemos hoy en día. En este último apartado, nos detenemos en dos aspectos que pueden dar forma a la evaluación que veremos en los próximos años: los nuevos desarrollos metodológicos y la implicación de las administraciones públicas.

Las técnicas de evaluación evolucionan rápidamente, y es importante incorporar en la docencia algunos de los desarrollos que se han dado en los últimos años. Siguiendo a Gërkhani y Miller (2022), destacamos dos avances principales: (1) la combinación de diferentes técnicas de investigación social; y (2) la incorporación de métodos y técnicas procedentes de fuera de las ciencias sociales. En el primer caso, destaca la inclusión de diseños experimentales en encuestas representativas. En muchos casos las agencias financiadoras de la evaluación están interesadas en que los resultados sean extrapolables a una población dada, pero, al mismo tiempo, no quieren renunciar a la fortaleza metodológica que supone la identificación experimental. Para obtener lo mejor de estas dos metodologías, tenemos los denominados «experimentos en encuesta» o «encuestas factoriales» de las que hemos hablado anteriormente. Pero la triangulación metodológica no debe detenerse en este instrumento y en muchas ocasiones querremos explorar métodos de «validación cruzada», es decir, afrontar la misma evaluación mediante diferentes métodos para, posteriormente contrastar y complementar los resultados obtenidos. Por ejemplo, en algunas ocasiones querremos complementar los experimentos de campo típicos de la evaluación de políticas con encuestas o entrevistas a la misma población para incidir sobre componentes más subjetivos de la evaluación. Si los experimentos nos informan acerca de cambios en los comportamientos de los participantes, las encuestas nos pueden decir algo acerca de las creencias y motivaciones que se hallan detrás de estos comportamientos. En cuanto a la incorporación de métodos y técnicas desde fuera de las ciencias sociales, cada vez es más frecuente

el interés en los factores biológicos que se encuentran detrás del comportamiento social. En este sentido, la investigación social también está incorporando técnicas de neuro-imagen o medidas psico-fisiológicas, como la conductividad de la piel o el *eye-tracking*. Un curso de evaluación experimental de políticas públicas generalmente no tiene espacio para profundizar en estas técnicas, pero los estudiantes deberían ser conscientes de que la investigación de frontera ya está incorporando técnicas tradicionalmente ajenas a la investigación social.

Por otro lado, a lo largo del artículo hemos puesto el foco en el papel fundamental de las administraciones públicas a la hora de hacer evaluación experimental. Para poder evaluar una política usando un *experimento controlado aleatorizado* (RCT), es necesario incluir desde el diseño la asignación aleatoria de participantes. Esto no es posible sin la complicidad y el compromiso de las instituciones. Además, hemos presentado otros tipos de experimentos que nos informan sobre los mecanismos y las decisiones subyacentes en una política pública, y que permiten a la administración decidir mejor qué políticas implementar y cómo hacerlo. En los últimos años hemos visto ejemplos de incorporación del análisis experimental a las políticas públicas que no habrían sido posibles sin un reconocimiento institucional de la importancia de evaluar el trabajo que se hace desde las administraciones. Sin embargo, para continuar por esta senda, es necesario un compromiso institucional completo, no sólo con incluir los elementos necesarios para la evaluación a la hora de diseñar políticas, sino también con actuar de acuerdo al resultado de dichas evaluaciones. Para que una evaluación se pueda desarrollar correctamente, debe hacerse sabiendo que el resultado será transparente y útil. Esto requiere de una reorientación de los modelos de gobernanza en las instituciones que, si bien no es sencilla, creemos que es posible, puesto que ya está en marcha.

## 7. CONCLUSIÓN

Aunque no hay un documento donde se afirme que la evaluación experimental es preferente, lo cierto es que la mayoría de instituciones internacionales, y lo empezamos a ver también en el contexto español, recomiendan el uso de la evaluación experimental. Este tipo de evaluaciones serán, cada vez más, una pieza clave en las políticas del futuro, y es por ello que deben incluirse necesariamente en cualquier programa de enseñanza sobre políticas públicas. A lo largo del artículo hemos puesto el énfasis en qué elementos son necesarios para que una evaluación de ese tipo pueda funcionar. Desde poner el foco en la idea de control, hasta repasar los tipos de diseños experimentales y los problemas de validez y éticos que suelen acompañar a las reflexiones metodológicas sobre la experimentación. Pero no podemos olvidarnos de los límites de esta aproximación.

El primero de ellos es especialmente grave y lo queremos llamar el «fetichismo de la experimentación». Cuando las instituciones financiadoras de las políticas muestran una predilección por la evaluación experimental, las agencias encargadas de la evaluación pueden hacer un uso superficial de la evaluación experimental, por ejemplo, diseñando varios tratamientos o intervenciones que no estén informados por ninguna teoría o experiencia anterior. La evaluación experimental cobra todo el sentido precisamente cuando tenemos hipótesis claras que queremos testar. La simple comparación de dos grupos que participan en intervenciones distintas apenas nos proporciona información relevante sobre la política.

En segundo lugar, las evaluaciones experimentales de impacto nos pueden ofrecer buena información acerca del cambio en variables objetivas, como el comportamiento observado de los participantes, pero no suelen ofrecer tan buena información sobre las motivaciones, creencias u otras variables subjetivas. En este caso es fundamental combinar la evaluación experimental con otras técnicas que permitan contemplar el impacto subjetivo de las políticas.

Además, en el artículo hemos repasado las principales consideraciones éticas relativas a la investigación experimental. Sin embargo, la ética no solo atañe al diseño específico de una evaluación, sino también a sus resultados y al uso que las administraciones hacen de ellos. Puede que se considere que ciertas intervenciones no deben ser evaluadas o que la ciudadanía prefiera una intervención diferente a la que ha sido favorablemente evaluada. La evaluación experimental es una herramienta muy útil para entender mejor cómo funcionan nuestras políticas, pero los gestores públicos serán en última instancia quienes deciden sobre las políticas a implementar.

Por último, la evaluación experimental está llena de obstáculos, como otros tipos de metodologías de investigación social, que los estudiantes e investigadores en el campo de las políticas públicas deben conocer. A lo largo del artículo hemos intentado repasar algunas de ellas, especialmente las relacionadas con los

sesgos de los participantes en la evaluación o las limitaciones prácticas y éticas al control experimental. Aun así, consideramos que la enseñanza de los métodos experimentales puede ser de gran utilidad para que los estudiantes reflexionen sobre la diferencia entre el diseño óptimo de una política pública y los condicionantes de la misma en la práctica.

## REFERENCIAS BIBLIOGRÁFICAS

- Akerlof, G., Oliver, A. y Sunstein, C. (2021). Editorial Introduction: Field experiments and public policy. *Behavioural Public Policy*, 5(1), número especial. <https://www.cambridge.org/core/journals/behavioural-public-policy/issue/8DBD9DA8775493E94495CFCC0796930D>
- Al-Ubaydli, O., Lee, M., List, J., Mackevicius, C., y Suskind, D. (2021). How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. *Behavioural Public Policy*, 5(1), 2-49. <http://doi.org/10.1017/bpp.2020.17>
- Al-Ubaydli, O. y List, J. A. (2015). On the generalizability of experimental results in economics. En G. Frechette y A. Schotter (eds.), *Handbook of Experimental Economic Methodology* (pp. 420-462). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195328325.003.0022>
- Angrist, J. D. y Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. <https://doi.org/10.1515/9781400829828>
- Baldassarri, D. y Abascal, M. (2017). Field experiments across the social sciences. *Annual Review of Sociology*, 43(1), 41-73. <https://doi.org/10.1146/annurev-soc-073014-112445>
- Barrera, D., Gërkhani, K., Kittel, B., Miller, L. y Wolbring, T. (2023). *Experimental Sociology: An Outline of a Scientific Field*. Cambridge University Press.
- Campbell, D. T., y Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Rand McNally & Company. <https://www.jameslindlibrary.org/campbell-dt-stanley-jc-1963/>
- Caparrós, A., Cerdá, E., Ovando, P. y Campos, P. (2007, 6 de abril). *Carbon Sequestration with Reforestations and Biodiversity-Scenic Values* [FEEM Working Paper 28]. Fondazione Eni Enrico Mattei. <https://services.bepress.com/feem/paper31/>
- Cook, T. D., Y Campbell, D. T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin.
- Deaton, A. y Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, (210), 2-21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- Dinas, E., Matakos, K., Xefferis, D., y Hangartner, D. (2019). Waking Up the Golden Dawn: does exposure to the refugee crisis increase support for extreme-right parties? *Political analysis*, 27(2), 244-254. <https://cadmus.eui.eu/handle/1814/61887>
- Dunning, T. (2012). *Natural Experiments in the Social Sciences. A Design-based Approach*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139084444>
- Falk, A. y Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535-538. <https://doi.org/10.1126/science.1168244>
- Gereke, J. y Gërkhani, K. (2019). Experimental Economics and Experimental Sociology. En J. H. Hamilton, A. Dixit, S. Edwards y K. Judd (eds.), *Oxford Research Encyclopedia, Economics and Finance*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190625979.013.462>
- Gerber, A. S., y Green, D. P. (2012). *Field experiments. Design, analysis, and interpretation*. W. W. Norton.
- Gertler, P. (2004). Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment. *American Economic Review*, 94(2), 336-341. <https://www.aeaweb.org/articles?id=10.1257/0002828041302109>
- Gertler, P., Martínez, S., Premand, P., Rawlings, L. B. y Vermeersch, C. M. J. (2016). *La Evaluación de Impacto en la Práctica* (2.ª Edición). Inter-American Development Bank – World Bank. <https://openknowledge.worldbank.org/handle/10986/25030>
- Gërkhani, K. (2020). Status Ranking and Gender Inequality: A Cross-Country Experimental Comparison. *Research in Social Stratification and Mobility*, (65), artículo 100474. <https://doi.org/10.1016/j.rssm.2020.100474>
- Gërkhani, K., Brandts, J. y Schram, A. (2021, septiembre). *Competition and gender inequality: A comprehensive analysis of effects and mechanisms* [BSE Working Paper 1292]. Barcelona School of Economics. <https://bse.eu/research/working-papers/competition-and-gender-inequality-comprehensive-analysis-effects-and>
- Gërkhani, K. & Miller, L. (2022). Experimental Sociology. En K. Gërkhani, N. D. de Graaf y W. Raub (eds.), *Handbook of Sociological Science. Contributions to Rigorous Sociology* (pp. 309-323), Edward Elgar Publishing.
- Glennerster, R. y Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton University Press.
- Harrison, G. W. y List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009-1055. <https://doi.org/10.1257/0022051043004577>

- Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, 19(3), 397-401. [https://doi.org/10.1016/s0167-4870\(98\)00013-0](https://doi.org/10.1016/s0167-4870(98)00013-0)
- Jiménez-Buedo, M. y Miller, L. (2010). Why a Trade-Off? The Relationship Between the Internal and External Validity of Experiments. *THEORIA, An International Journal for Theory, History and Foundations of Science*, 25(3), 301-321. <https://ojs.ehu.es/index.php/THEORIA/article/view/779>
- John, P. (2017). *Field Experiments in Political Science and Public Policy. Practical Lessons in Design and Delivery*. Routledge. <https://doi.org/10.4324/9781315773025>
- Miller, L. (2006). Experimentos de orientación teórica: una discusión metodológica, *Empiria* (12), 89-110.
- Morton, R. B. y Williams, K. C. (2010). *Experimental Political Science and the Study of Causality. From Nature to the Lab*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511762888>
- Mutz, D. C. (2011). *Population-Based survey experiments*. Princeton University Press. <https://doi.org/10.23943/princeton/9780691144511.001.0001>
- Nakamura, E., Sigurdsson, J. y Steinsson, J. (2022) The Gift of Moving: Intergenerational Consequences of a Mobility Shock. *The Review of Economic Studies*, 89(3), 1557-1592. <https://doi.org/10.1093/restud/rdab062>
- Oakley A. (1998). Experimentation and social interventions: a forgotten but important history. *British Medical Journal*, 317(7167), 1239-1242. <https://doi.org/10.1136/bmj.317.7167.1239>
- Parker, S. W. y Teruel, G. M. (2005). Randomization and Social Program Evaluation: The Case of Progreso. *The ANNALS of the American Academy of Political and Social Science*, 599(1), 199-219. <https://doi.org/10.1177/0002716205274515>
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies. *Journal of Educational Psychology* 66(5), 688-701. <https://doi.org/10.1037/h0037350>
- Salganik, M. J. (2019). *Bit by Bit: Social research in the Digital Age*. Princeton University Press.
- Schram, A., Brandts, J. y Gërkhani, K. (2019). Status ranking: a hidden channel to gender inequality under competition. *Experimental Economics*, (22), 396-418. <https://doi.org/10.1007/s10683-018-9563-6>
- Shadish, W. R., Cook, T. D. y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.
- Webster, M. y Sell, J. (2014). *Laboratory Experiments in the Social Sciences*. Elsevier. <https://doi.org/10.1016/C2011-0-07562-2>
- Weiss, C. H. y Birckmayer, J. (2009) Social Experimentation for Public Policy. En R. Goodin, M. Moran y M. Rein (eds.), *The Oxford Handbook of Public Policy* (pp. 806-830). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199548453.003.0039>