



El abuso de la ciencia en la política

por **Benito Arruñada**

La pretensión de someter las decisiones políticas, empresariales y organizativas a la supuesta evidencia científica es fuente común de errores y oportunismos. Una triada de pautas puede ayudar a mejorar el uso de la ciencia y la información en la toma de decisiones públicas.

Hace algún tiempo, una de mis tribunas de prensa sobre educación fue criticada por incluir demasiadas anécdotas. La crítica era valiosa porque abría el debate de qué información debemos usar para las decisiones sociales, y porque este asunto de los distintos tipos de evidencia y, en general, del papel que puede representar la ciencia en la toma de decisiones políticas y organizativas merece un análisis detenido, para aportar algo de claridad y evitar que, paradójicamente, contribuya al oscurantismo.¹

¹ Supongo aquí resuelto un asunto crucial: antes de evaluar, hay que saber *qué producimos*. Por ejemplo, si, aunque digan lo contrario, para muchos ciudadanos la enseñanza pública es un producto de consumo, no es sensato tratarlo como si fuera una inversión. En puridad, una medida completa para la toma de decisiones debería incluir todos los *inputs* (incluido el esfuerzo del

Información para la toma de decisiones

Aunque a costa de crear otras dificultades, el problema nos viene resuelto cuando decidimos a través de ese planificador descentralizado que es la economía de mercado, pues este opera de forma automática, sin necesidad de gestión consciente. Gracias a que el derecho de propiedad incentiva a los decisores y el sistema de precios los informa del valor de bienes y servicios, todo mercado, automáticamente, asigna los recursos entre usos

estudiante) y *outputs* (aptitud y actitud, pero también ideología) del proceso educativo. El olvido sistemático de ciertas dimensiones revela qué atributos se desprecian o, quizá, se consideran impresentables (esfuerzo, actitudes, ideología).

alternativos, evalúa a sus propietarios y los premia cuando aciertan en sus decisiones.

En cambio, tanto en el seno de las empresas como en los Estados y en sus administraciones, esas tareas de asignación, evaluación y compensación requieren mecanismos artificiales y quedan sujetas a todo tipo de manipulaciones. Dado que no hay precios, hemos de producir y transmitir información para conocer la demanda y organizar la oferta, tomar decisiones y evaluarlas.² Además, como la mayoría de los decisores no ostentan derechos de propiedad sino que deciden por delegación, aparecen graves asimetrías informativas, y los conflictos de intereses se vuelven crónicos. En consecuencia, esos sistemas artificiales de planificación no son solo solución sino también parte esencial del problema organizativo. De hecho, ello supone una limitación primordial a toda pretensión de resolver los fallos del mercado mediante intervenciones y burocracias públicas, pues estas los acaban padeciendo, a menudo en versión agravada. Por este motivo, es esencial mantener en esta materia una perspectiva coasiana en cuanto al imperativo de *comparar realidades imperfectas*,³ sin tomar nunca como referencia idealismos falaces de uno u otro signo (ni mercados perfectos ni estados benevolentes).

Evaluación decorativa

De entrada, a menudo, si dentro de una organización no se decide no es por falta de información, sino porque los derechos de decisión están mal ubicados: quien tiene autoridad no sabe y quien sabe no tiene autoridad o no le conviene tomar la decisión correcta.

En esos casos, no es tan necesario producir información adicional como reasignar los recursos o los derechos de decisión, y cambiar los incentivos de los decisores. Sin embargo, lo frecuente es que se dediquen recursos a tareas de evaluación decorativa, consistentes en producir información sobre el rendimiento a sabiendas de que los resultados de esa evaluación no darán lugar a decisiones de compensación o reasignación de recursos.

Con frecuencia, lo que se persigue es dedicar recursos para evaluar con el único fin de dar la impresión de que se gestiona y, de paso, contratar consultores y crear gabinetes de planificación en los que colocar amiguetes. Ha sido este un fenómeno común en las reformas características del *new public management*, incluidos los “mercados internos” del tipo del National Health Service británico. Lo único que

aseguran estas iniciativas es la creación de un *Gosplan* planificador de utilidad más bien discutible.

Por ejemplo, volviendo a la educación española, tendría poco sentido desarrollar un sistema más complejo para gestionar la enseñanza concertada (un modelo de hecho cercano a los propios de los mercados internos) cuando, pese a que ha aumentado la demanda de los centros concertados y disminuido la de los públicos, la mayoría de los gobiernos regionales ha restringido la oferta de los concertados. Esto es, serviría de poco conocer mejor el rendimiento de ambos tipos de centros y su demanda cuando, en nuestras decisiones políticas, ya nos hemos esforzado en contradecir las preferencias de la ciudadanía: una sociedad no necesita saber qué quiere el ciudadano si ya ha demostrado que, en todo caso, esa demanda ciudadana no le importa. La demanda se conoce y, a menudo, la acumulación de datos solo busca, en el fondo, negarle autoridad para centralizar las decisiones en los aparatos administrativos, sustrayendo la capacidad de decidir del ciudadano.

El asunto es aún más claro en el ámbito universitario, pues tampoco ha servido de gran cosa producir estadísticas de empleo y salarios de los graduados de cada centro y carrera universitaria, ya que, pasmosamente, esos datos permanecen secretos y no se dan a conocer al público, salvo en Cataluña, donde, con todo, se divulgan con escasa publicidad y de forma tal que resulta difícil comparar los ingresos promedio que logran los graduados de cada centro y cada carrera.⁴

Evaluación manipulativa

Si la evaluación decorativa es mala, resultan aún más perniciosos los fenómenos de evaluación manipulativa, en los que se produce información (a la que se suele presentar como “evidencia”) para justificar políticas interesadas y a menudo contrarias al bien común. Sucede así cuando se recoge información selectiva de algunas variables o se las mide con un sesgo favorable, de modo que queden bien y se reivindicuen las políticas previamente elegidas.

Algunas de las estadísticas y encuestas de nuestra enseñanza padecen este síndrome en grado notable: por ejemplo, las pruebas PISA realizadas en Cataluña pretenden evaluar el conocimiento del español efectuando solo exámenes en catalán. También cuando se reconfigura y canaliza el uso de las categorías jurídicas para que el propio aparato estadístico recoja cambios cosméticos que reflejen una evolución falsamente positiva, como ha sucedido recientemente

2 Véase, por ejemplo, Kenneth J. Arrow, “Control in large organizations”, *Management Science*, 1964, 10(3), pp. 397-408, y Michael J. Jensen y William H. Meckling, “Specific knowledge and divisional performance measurement”, *Journal of Applied Corporate Finance*, 2009, 21(2), pp. 49-57.

3 La clave sigue estando en las páginas finales, tan poco leídas y nada atendidas, de Ronald H. Coase, “The problem of social cost”, *Journal of law and economics*, 1960, 3, pp. 1-44.

4 Agencia para la Calidad del Sistema Universitario de Catalunya, “Tota la informació sobre la qualitat dels estudis oficials universitaris de Catalunya”, 2023, disponible en: <https://estudis.aqu.cat/euc/ca/Comu/Inici>.

5 El que en las empresas no se suele hablar de “evidencia” sino de “información” sugiere que el uso del término “evidencia” en las decisiones públicas sirve funciones meramente retóricas.

con la transformación puramente formal de los antiguos contratos temporales en “fijos discontinuos”.

Por este camino, las políticas que pretenden basarse en la evidencia —la célebre *evidence-based policy*— acaban transformándose en una recogida de evidencia orientada a justificar políticas previamente elegidas por todo tipo de motivos e intereses más o menos espurios. Terminan así en la fabricación de mera *policy-based evidence*.

Políticas para el postuero

Las peores mixtificaciones se producen cuando se cambia de política para manipular los indicadores de rendimiento. Es lo que ha ocurrido cuando nuestros gobiernos, con la complicidad de los centros académicos, redujeron los estándares de rendimiento al suprimir los exámenes independientes de los centros, ablandar los niveles de exigencia de los exámenes, permitir pasar de curso sin haber aprobado el anterior y otorgar titulaciones con asignaturas suspensas o aprobadas “por compensación”. En buena medida, estas políticas buscaban regalar titulaciones para reducir artificialmente los indicadores de abandono escolar y así maquillar los malos resultados obtenidos por nuestro país en las comparaciones internacionales. Se produce un efecto similar a menor escala cuando los estudiantes adaptan su estudio no para aprender sino para aprobar los exámenes o cuando, de modo similar, los propios profesores adaptan su docencia con el único fin de que sus estudiantes superen exámenes externos independientes.

Se trata de una limitación insalvable, pues se cumple la ley de Campbell:⁶ cuanto más se usa una información para tomar decisiones, reasignando recursos y recompensando a los participantes, más incentivos se ligan a esos indicadores, datos o evidencias, y, como consecuencia, estos tienden a distorsionarse y perder valor informativo. Si a los colegios y profesores se les premia, como sucede en España, por el porcentaje de aprobados, el hecho de aprobar acaba dejando de ser informativo de la preparación del estudiante.

Por estos motivos, la recogida de información cuantitativa debe manejarse con sumo cuidado para evitar la manipulación de los propios indicadores. La gestión de incentivos empresariales aprendió hace muchos años, tras fracasar el *management by the numbers* introducido con la generalización de los ordenadores en las grandes empresas, a no interpretar de forma automática las desviaciones respecto a estándares o promedios: una desviación aparentemente positiva es solo una señal de alarma, que como tal debe ser investigada, pero nunca debe tomarse sin más como una medida de rendimiento, sin descartar antes que obedezca a una

manipulación de los indicadores, que a menudo perjudica el interés social.⁷

Por los mismos motivos, al evaluar el rendimiento de personas y organizaciones conviene combinar métodos y criterios objetivos y subjetivos. Ambos sufren deficiencias, pero en buena medida son complementarios, es labor del buen directivo encontrar y potenciar esas complementariedades.⁸ En términos de modas gerenciales, el *management by the numbers* pronto pasó a ser complementado en empresas como la mejor Hewlett-Packard de los años 1970 por el *management by wandering around*.⁹

Se comprende en esta línea el doble papel crítico que representan las anécdotas como fuente de información. Por un lado, inician la evaluación cualitativa del proceso productivo, la cual desnuda la ficción de unos indicadores cuantitativos que más de una vez han sido manipulados torticeramente. Por otro lado, al mostrarnos realidades olvidadas por el cientificismo al uso, desvelan su impostura. No son toda la verdad, pero tampoco lo pretenden; y no solo muestran las limitaciones de la verdad oficial, sino que completan la información del decisor.

En el terreno educativo, solo resta señalar que la propia OCDE también tiene sus intereses cortoplacistas: más que las reformas de verdad, que suelen ser impopulares y dar sus frutos a largo plazo, a sus burócratas les interesa parecer importantes en el presente, motivo por el cual comunican los resultados del estudio PISA de modo que se hable de ellos en la prensa y que los gobiernos los sigan, aunque sea al coste de simplificar sus análisis, presentarlos de forma emocional (por ejemplo, como una liga deportiva) y enturbiar sus recomendaciones.¹⁰

7 Una lección básica de la contabilidad gerencial. Véase, por ejemplo, Robert Kaplan y Anthony Atkinson, *Advanced management accounting*, Prentice-Hall, Englewood Cliffs, 2013.

8 Un trabajo clásico sobre la función de la evaluación subjetiva es el de George Baker, Robert Gibbons y Kevin J. Murphy, “Subjective performance measures in optimal incentive contracts”, *Quarterly Journal of Economics*, 1994, 109(4), pp. 1125-1156. Para una introducción a esta problemática, véase Benito Arruñada, *Empresa, mercado e instituciones*, Thomson-Reuters, Cizur Menor, 2013, pp. 263-311.

9 Popularizado más tarde por Tom Peters y Robert H. Waterman Jr., *In search of excellence: Lessons from America's best-run companies*, 1982, Harper & Row, Nueva York, así como por su versión japonesa del más formal *gembu walk*, desarrollado por Toyota como parte integral de su filosofía de *lean management* y *just-in-time manufacturing*.

10 La OCDE sufre aquí el mismo síndrome que sufría el Banco Mundial respecto al proyecto “Doing business”, hoy caído en desgracia: sabía que sus métodos eran incorrectos, que había corrupción y que animaba reformas cosméticas e incluso contraproducentes, pero todo se toleraba mientras que la prensa financiera se lo tomara en serio e informara de sus informes anuales. Le hacía parecer importante en un momento en el que la relevancia del propio banco estaba en entredicho. He analizado este caso en Benito Arruñada, “Pitfalls to avoid when measuring the institutional environment: Is doing business damaging business?”, *Journal of comparative economics*, 35(4), 2007, pp. 729-747, y “How doing business jeopardizes institutional reform?”, *European business organization law review*, 2009, 10(4), pp. 555-574.

6 Donald T. Campbell, “Assessing the impact of planned social change,” *Evaluation and program planning*, 2(1), 1979, pp. 67-90. (The Public Affairs Center, Dartmouth College, Paper 8, Occasional paper series, 1976, <https://shorturl.at/kNSWY>).

La trampa del falso rigor científicista

En el fondo, esta manipulación o *gaming* de indicadores no es sino una manifestación más del principio de incertidumbre de Heisenberg en el terreno social. Unida a la mayor complejidad intrínseca de los problemas sociales,¹¹ hace que sean menos susceptibles al tipo de tratamiento que en muchos dominios de las ciencias físicas se considera científico. Por su menor capacidad de predicción, el papel de las ciencias sociales ha de ser por ello más modesto y respetuoso con soluciones heredadas, que son fruto de la evolución cultural y cuya estructura y lógica interna desconocemos.

Sin embargo, mucho científico social tiende a ocultar esas dificultades.¹² La última moda consiste en exagerar el rigor de la evaluación a riesgo de trivializarla o centrarla en aspectos secundarios, olvidando de paso asuntos centrales pero incómodos para el poder como son, en el caso de la educación, el tamaño de las clases, la selección del profesorado, la relajación de la exigencia académica o los efectos reales en la equidad.

Esta reciente obsesión de parte de las ciencias sociales por una identificación causal rigurosa ha elevado los criterios para aceptar un resultado como verosímil, pero lleva a investigar tan solo aquellas cuestiones para las que existan o se puedan producir datos apropiados, en una versión no por intelectual menos ridícula de quien busca sus llaves extraviadas allí donde hay luz y no donde cree que pueden haber caído.¹³

Poco importa que el supuesto rigor deje a menudo bastante que desear, por ignorar el carácter sistémico de la realidad. Por ejemplo, se suele marginar el hecho de que la presencia de errores de medida —en especial, en las variables independientes— introduce un grado de endogeneidad suficiente para cuestionar los resultados. También se suelen minimizar las dificultades que existen en los experimentos sociales para efectuar pruebas en materias de cierta importancia, no solo porque las opciones en juego son demasiado valiosas, sino porque cuanto más valiosas sean, más difícil es asegurar verdadera aleatoriedad.

En todo caso, el requisito es que algún fenómeno aleatorio proporcione una excusa aceptable para justificar la estrategia de identificación causal. Con ello, se limita drásticamente el tipo de problema en estudio: ya no se toma como objeto de estudio el más relevante, sino aquel del que se pueda argumentar una metodología de identificación causal rigurosa, ya sea por la preexistencia de datos sometidos

a una circunstancia aleatoria o por la posibilidad de desarrollar expresamente un experimento aleatorio.¹⁴

Como consecuencia, las conclusiones son difícilmente extrapolables y sufren un serio déficit de “validez externa”: cuanto más rigurosos son los resultados, más se refieren a unas circunstancias muy concretas. Paradójicamente, esta deficiencia se emplea con frecuencia para justificar políticas que son insensatas desde un punto de vista general, con el argumento de que en algunas circunstancias remotas pudieran tener efectos positivos. Para ello, se recomienda experimentar con diversas regulaciones y construir complejos sistemas informativos con la pretensión de recoger información detallada de sus efectos. Resultados relativamente rigurosos pero que pueden obedecer a particularidades del caso y carecen de validez externa sirven así para justificar propuestas insensatas y costosas porque “todo es posible” en algunos supuestos y, en el peor de los casos, su adopción siempre permitirá “medir mejor” sus efectos.¹⁵

Por desgracia, dada la reiteración de las limitaciones, la información adicional sufrirá similares problemas y será igual de irrelevante. La acumulación de evidencia rigurosa es inútil si carece de *utilidad decisional* porque se refiere solo a algunos de los factores de decisión, y, como consecuencia, distintos factores se conocen con un grado de precisión muy diferente. Por ejemplo, si una decisión se basa en comparar el coste A y el beneficio B, es inútil conocer A con exactitud si se desconoce hasta el orden de magnitud de B. Desde el punto de vista del bienestar social, sería generalmente preferible conocer tanto A como B de modo imperfecto. La cuestión clave es cómo asignar recursos eficientemente entre la producción de información sobre A y sobre B.¹⁶ Lo socialmente óptimo es de forma equilibrada; pero, en muchos terrenos, los incentivos imperantes en las ciencias sociales llevan a una asignación desequilibrada. Esta depende en buena medida de la competencia en los “mercados” de ideas: es más equilibrada en una litigación, pero tiende a serlo mucho menos en el ámbito estrictamente académico, donde reina la mera acumulación de conocimientos muy precisos sobre A es o será algún día socialmente valiosa pese a no saber casi nada sobre B.

Todas estas debilidades quedan disimuladas si la calidad científica viene respaldada por los criterios de

14 Christopher A. Sims, “But economics is not an experimental science”, *Journal of Economic Perspectives*, 2010, 24(2), pp. 59-68.

15 Este tipo de excusa científicista ha servido, por ejemplo, para dar cierta respetabilidad a la resurrección de regulaciones sobre las que, durante décadas, había existido un cierto consenso en considerarlas generalmente nocivas. Es el caso de las relativas al aumento del salario mínimo o a la congelación de los alquileres, con consecuencias que están siendo socialmente nefastas.

16 La pretensión de identificar causalidades sin validez externa y, por tanto, de escasa relevancia, confiando en alcanzar la verdad completa por acumulación futura de las causas, se contraponen aquí a la modestia de un proceso más *popperiano* de falsación de conjeturas, basadas en observar correlaciones refutables sobre los factores relevantes para las decisiones.

11 Friedrich A. Hayek, “The theory of complex phenomena”, en M. Bunge (ed.), *The critical approach to science and philosophy*, 1964, Nueva York, Free Press.

12 Un punto que Friedrich A. Hayek subrayó en su discurso de recepción del Premio Nobel: “The pretence of knowledge”, *American economic review*, 1989, 79(6), pp. 3-7.

13 George A. Akerlof, “Sins of omission and the practice of economics”, *Journal of economic literature*, 2020, 58(2), pp. 405-418.

respetabilidad de la correspondiente disciplina especializada,¹⁷ como tal vez sucede hoy en economía tras el momentáneo éxito de la *credibility* o *identification revolution*. Sea cual sea su validez externa, los descubridores de esas anomalías y excepciones ganan pedigrí académico y son invitados a informar todo tipo de decisiones sobre las que a menudo tienen, inevitablemente, un conocimiento limitado: por lo especializado de sus investigaciones, suelen dominar con precisión lo trivial al precio de ignorar —ni siquiera de forma aproximada— lo más relevante y, sobre todo, de carecer de la perspectiva y de la *expertise* necesarias para identificar prioridades.

Cómo usar y no usar la evidencia en las decisiones

A modo de resumen, concluiré proponiendo tres pautas elementales para el uso de la evidencia científica en las decisiones sociales, relativas a la coordinación, el equilibrio y la complementariedad.

1. *Coordinación*. La producción de información es costosa y su coste ha de ser inferior a los beneficios que proporciona. Ello implica que se adecue a las decisiones que se quiera y se puedan tomar. No procede producir más información, causando derroche, distracción y conflicto, si la información ya disponible es suficiente para decidir y, sin embargo, no se la está utilizando.

2. *Equilibrio*. Los sistemas sociales suelen ser más complejos que los físicos y tienen una lógica interna cuya estructura generalmente desconocemos, por lo que el decisor debe moderar sus pretensiones y ser prudente al modificarla. Para considerar esa complejidad con eficacia, la información para la toma de decisiones ha de ofrecer una calidad similar en todos los factores esenciales, lo que implica rechazar el rigorismo parcial por costoso e irrelevante.

3. *Complementariedad*. Buena parte de la complejidad de los sistemas sociales obedece a que los seres humanos responden de manera creativa e inesperada a todo sistema de incentivos. El uso de la información ha de tener en cuenta los efectos que la propia recogida de información ejerce en los agentes, lo que hace necesario emplear de forma complementaria tanto información cuantitativa como cualitativa, así como objetiva y subjetiva. ~

BENITO ARRUÑADA es catedrático de organización de empresas de la Universidad Pompeu Fabra.

¹⁷ Bien entendido que la especialización disciplinar ofrece obvias ventajas en términos de menor coste de comunicación y mayor eficacia del control de calidad, como ya analizaron en su día Michael Polanyi (“The republic of science: Its political and economic theory”, *Minerva*, 1962, 1(1), pp. 54-73) y Philip Kitcher (*The advancement of science: Science without legend, objectivity without illusions*, Oxford y Nueva York, Oxford University Press, 1993). Sin embargo, la cuestión difícil para las ciencias sociales es si esas ventajas claudican o no ante los imperativos del “mercado” político de las ideas, las creencias y las excusas.

