

La evaluación de «ET CHOICES»¹

Robert D. Behn *

Este artículo es un excelente ejemplo de análisis de una política. Concretamente se analiza la evaluación de un programa del Departamento de Bienestar Público del Estado de Massachusetts. El artículo puede interesar tanto a los gestores o investigadores de políticas sociales similares como a los que tengan interés en la discusión de algunos problemas que suele presentar la evaluación de políticas. El argumento central es que no se puede evaluar en este caso con un experimento social, como insistentemente pedían algunos responsables políticos, dadas las características del programa. En último término, el autor acepta que no es posible determinar de forma definitiva en qué medida el programa ayuda a sus beneficiarios a salir de su situación de pobreza. Sin embargo, al mismo tiempo, entiende que los estudios realizados tanto por la misma administración como por institutos independientes, aportan suficiente evidencia empírica para, desde la perspectiva de gestor público —que es distinta a la del científico-evaluador—, determinar el valor del programa. Xavier BALLART.

* * *

En enero de 1986, el Departamento de Bienestar Público del Estado de Massachusetts (DB) publicó la Evaluación del Programa de Empleo y Formación «ET CHOICES», preparada por la Oficina de Investigación, Planificación y Evaluación del Departamento (RP&E). La evaluación estaba pensada para responder a «cuestiones sobre el número y las características de los participantes en el programa». Por ejemplo, el informe afirmaba que, «si bien los participantes eran representativos de la población AFDC², estaban más a punto para conseguir un empleo». En el mismo sentido, más participantes tenían «experiencia de trabajo previa» (87 por ciento) en comparación con los beneficiarios de AFDC en general (73 por ciento)³.

RP&E también examinó «el impacto del programa en las experiencias de empleo y en la dependencia de los programas de asistencia social de los participantes, así como los efectos en las subvenciones de asistencia social». Por ejemplo, el informe afir-

maba: «la media de ingresos por hora trabajada por empleos a tiempo completo conseguidos a través de los servicios de formación y colocación de ET era de \$5.11, más del doble de los ingresos que se obtienen con una prestación del programa AFDC y 53% por encima del salario mínimo por hora trabajada». También indicaba que la «media de ahorro para el Estado por un trabajo conseguido a través de los servicios de formación y colocación era de \$219»⁴.

La evaluación del DB sugería tres maneras en las que se podía mejorar el programa: (1) reduciendo el *gap* salarial entre blancos y minorías, (2) buscando trabajos mejor pagados, y (3) poniendo un mayor énfasis en formación antes que en la colocación directa:

Entre los diversos componentes del programa, los componentes de formación en nuevas habilidades y experiencia de trabajo real tienen las proporciones más altas de empleo y terminación de la dependencia del Estado. Ello demuestra el valor de mejorar las habilidades vocacionales de los beneficiarios a través de un programa de formación intensivo como estrategia para reducir la dependencia de la asistencia social. Es especialmente significativo en ET porque los participantes en estos componentes estaban menos preparados para conseguir un empleo que los participantes en el componente de colocación directa.

Dedicando más recursos a programas más intensivos en formación puede, a corto plazo, reducirse el número de colocaciones, dado que los participantes en cursos de educación básica o formación en habilidades permanecen más tiempo en el programa. A largo plazo, sin embargo, ello resultaría en mejores trabajos para los participantes y mayores ahorros para el Estado. Si bien los programas de formación son más caros, también son más efectivos en ayudar a los beneficiarios a prepararse para / encontrar trabajos que les permitirán salir de / permanecer al margen de la asistencia social, si se compara con los servicios de colocación menos intensivos en desarrollo personal⁵.

Esta evaluación no incluía todos los datos y análisis de los que disponía el DB. RP&E constantemente recogía información

y la examinaba. A menudo, simplemente, resumía estas informaciones en gráficos para publicaciones o presentaciones. A veces, las utilizaba para informes internos. En algunas ocasiones, el personal encargado de escribir discursos para el Gobernador M. Dukakis citaba los datos. En otras ocasiones, el DB los utilizaba para explicar a la prensa y al público en general como estaba funcionando ET. Una vez al año, el DB utilizaba la información disponible para explicar su proyecto de presupuesto al legislativo.

Al mismo tiempo, la información recogida no podía mostrar lo que habría pasado en Massachusetts sin ET CHOICES. El DB quería saber cómo estaba funcionando el programa y cómo podía ser mejorado. Por ello, hizo un esfuerzo importante para recoger información que permitiera responder a estas cuestiones.

La estimación de la reducción de casos y del ahorro presupuestario

El DB quería demostrar que ET CHOICES estaba reduciendo el número de personas dependientes de las subvenciones de asistencia social y por tanto ahorrando dinero al Estado. Por ejemplo, en el apartado correspondiente a ET dentro del proyecto de presupuesto para 1987, el DB indicaba como desde enero de 1983 hasta julio de 1985, el número de casos de AFDC se había reducido en un 9.4% —la reducción más importante entre los doce Estados con mayor número de casos (de los doce, seis habían aumentado el número de casos)—. Durante este período, el nivel de desempleo se redujo en los doce Estados, cinco sin embargo, experimentaron disminuciones mayores que Massachusetts. “Una economía fuerte por sí misma no es suficiente para dar lugar a una disminución importante en el número de AFDC”, concluía el documento presupuestario ⁶. Se requiere, por tanto, una explicación adicional.

“En Massachusetts,” afirmaba el documento, “ET CHOICES se ha convertido en una de las principales causas de la reducción en el número de casos.” El presupuesto también incluía un gráfico en el que contrastaba el número de casos con el estimado sin ET. El documento afirmaba, “El número de casos AFDC en noviembre de 1985 sería de 95.945 comparado con los 85.411 si no fuera por ET” ⁷. El presupuesto, sin embargo, no incluía los cálculos por los que las 23.000 colocaciones del programa se traducían en una disminución de 10.500 en el número de casos. Tampoco explicaba el método empleado para obtener estas estimaciones.

En 1988, el proyecto presupuestario, en el apartado para el programa ET, incluía nuevos datos: “El número de casos AFDC ha disminuido un 4.2% —desde 88.414 en octubre de 1983 cuando ET empezó a 84.693 en diciembre de 1986— a pesar de la significativa presión que ha supuesto un incremento del 32% en los beneficios, un incremento del 35% derivado de los cambios en el nivel de elegibilidad para entrar en el programa y un 87% de incremento en el número de familias monoparentales en la última década. Al contrario, los casos AFDC en los doce Estados con mayor número de personas dependientes de asistencia social aumentaron, como media, en un 6% durante el mismo período. El DB estima que sin ET, el número de casos de AFDC hubiera sido en torno a unos 100.000” ⁸. Nuevamente, sin embargo, el DB no explicaba cómo estimaba el número de casos sin ET.

Pero el presupuesto de 1987 sí explicaba como el DB concluía que ET había ahorrado ya al Estado \$69 millones. El coste total de colocar 23.000 participantes había sido de \$71 millones (con un coste medio por colocación de unos \$3.100). El coste medio anual para un caso AFDC era de \$6.100. Por tanto, el total anual no incurrido en gastos para 23.000 casos era de \$140 millones. La diferencia entre los gastos en ET de \$71 millones y los gastos no incurridos en AFDC de \$140 millones eran los ahorros estimados de \$69 millones ⁹.

En agosto de 1987, la Fundación de Contribuyentes de Massachusetts ¹⁰ (“cuya misión es ayudar al Estado a funcionar de una manera más eficiente”) publicó su propia evaluación de ET incluyendo una estimación del ahorro presupuestario. Suponiendo que el 60% de las colocaciones resultara en casos cerrados y “que aquellos que dejan de recibir dinero público gracias a ET, hubieran permanecido en el programa durante otro año si no fuera por el programa,” la Fundación proyectó ahorros para un período de cinco años de \$150 millones hasta junio de 1988 ¹¹. Los ahorros anuales alcanzaban su máximo en 1986 y después se reducían debido al aumento en el coste medio por colocación a través de ET.

Al estimar el ahorro presupuestario, ambos, el DB y la Fundación suponían que ninguno de los beneficiarios que obtenía un empleo hubiera dejado AFDC sin ET. Por tanto, ambos cálculos sobreestimaban los ahorros producidos por ET. El cálculo del DB claramente sobreestimaba los ahorros al suponer que cada puesto de trabajo ET automáticamente producía el cierre del caso en AFDC (reduciendo a cero la prestación). Sin embargo, la misma evaluación del DB reflejaba que, durante los dos primeros años del programa, sólo un 45 por ciento de las colocaciones ET daban lugar a casos cerrados en AFDC. Por tanto, por lo menos durante este período, los cálculos de la Fundación de los Contribuyentes (que suponía que un 60 por ciento de las colocaciones ET resultaban en un caso cerrado) también sobreestimaban el ahorro presupuestario.

Ambos, el DB y la Fundación también suponían que los beneficiarios de AFDC que encontraban trabajo a través de ET sólo lograban no volver a depender de la asistencia social durante un año. Esta suposición puede haber infravalorado los ahorros debidos a ET. Si, como media, los antiguos beneficiarios de ET conseguían no volver a depender de la prestación pública por espacio de dos años, el ahorro real hubiera sido más del doble del ahorro estimado. Obviamente, los esfuerzos para estimar los ahorros presupuestarios de cualquier programa de asistencia social, formación y trabajo son sensibles a múltiples suposiciones.

Hay una manera más sencilla de ver si un programa es coste-eficaz. Se puede ignorar la cuestión de si un beneficiario hubiera obtenido el trabajo (o salido de la asistencia social), sin ET. En su lugar, el análisis se puede centrar en la siguiente cuestión: ¿Cuántos meses *antes* de que un participante hubiera salido de AFDC por sí mismo suele el participante medio de ET conseguir su independencia económica? Supongamos (como hace el presupuesto para 1987) que el coste medio de una colocación ET era de \$3.000 y que el coste medio anual para un caso AFDC era de \$6.000: luego, si como media, los participantes dejaban AFDC seis meses antes, el programa ahorra dinero al Estado a corto plazo. Si el coste de una colocación era de \$5.000 y el coste anual por AFDC era de \$8.700 (la estimación de la Fundación de Contribuyentes para 1987), el DB tenía que conseguir la independencia económica de los participantes siete meses antes para garantizar ahorros a corto plazo.

Si un programa de bienestar social, formación y empleo pasa este test de ahorros a corto plazo debido a la anticipación del cierre de los casos, el programa es claramente coste-eficaz. Este test, sin embargo, tendería a subestimar de dos maneras el ahorro presupuestario total para el programa. Primero, porque ignora los ahorros que resultan de un trabajo a tiempo parcial que produce una reducción en la prestación AFDC. Además, porque este test ignora algunos ahorros potenciales a largo plazo. Si, debido a su formación ET, algunos beneficiarios no vuelven a AFDC - o simplemente tardan más a volver a depender de las ayudas públicas - se producirían ahorros adicionales en el futuro. Por supuesto, estos ahorros adicionales aún son más difíciles de estimar.

Efectivamente, cualquier esfuerzo por estimar el ahorro presupuestario producido por ET CHOICES dramatiza la dificultad de evaluar cualquier programa público sin un grupo de control.

El diseño de un experimento para ET CHOICES

Supongamos que los líderes políticos de Massachusetts quisieran evaluar ET CHOICES con un experimento. El diseño de

este experimento tendría que reflejar las nueve características claves del programa:

1. ET debería ser voluntario. Nadie tendría que ser forzado a participar, ni tampoco se podrían reducir las prestaciones a aquellas personas que no quisieran participar.

2. Para garantizar que cada beneficiario de las ayudas públicas conoce ET y puede beneficiarse del programa, el DB debería dedicar recursos suficientes a la función de informar sobre el programa, convencer a sus beneficiarios que tienen la inteligencia necesaria para seguir un programa de formación, asegurarles que tienen el carácter necesario para ser económicamente independientes para, finalmente, seleccionarlos para su participación en el programa.

3. Este esfuerzo de *marketing* debería ser completado con otro similar dirigido a los empleados del DB para convencerles que los beneficiarios pueden ser ciudadanos productivos, y aún con otro dirigido a los posibles proveedores de trabajo para persuadirles en el sentido de que los beneficiarios de ayudas públicas quieren trabajar, además de ser perfectamente razonable desde la perspectiva empresarial contratar a estas personas.

4. ET debería extenderse por todo el Estado. Cada mujer dependiente de las ayudas públicas en el Estado debería tener la oportunidad de participar.

5. Una oficina de la agencia de colocación del Estado debería localizarse al lado de cada oficina del DB. De este modo, la provisión de servicios a los beneficiarios se produciría de una manera más directa y efectiva.

6. Una amplia variedad de servicios educativos, de formación para el empleo y de colocación deberían garantizar que cada participante ET encontrara un programa que pudiera satisfacer sus necesidades. Estos servicios no deberían ser distribuidos por el DB sino por organizaciones que tienen experiencia y se dedican a esta actividad. Ello garantizaría que (a) el programa utiliza a los que están mejor preparados para tareas de formación para el empleo y (b) estas organizaciones no reducen el número de servicios que ya están prestando. (Una cláusula en el contrato con estas organizaciones debe introducirse para asegurar que se mantiene el mismo nivel de esfuerzo).

7. Desgraciadamente, los promotores de ET no saben exactamente cómo va a funcionar el programa (ya que esto depende de la interacción incierta entre condiciones económicas locales, cultura política, características particulares de cada caso y del comportamiento individual y colectivo de los trabajadores sociales.) Por tanto, la eficacia de los distintos componentes del programa debería ser continuamente examinada y las lecciones e ideas generadas de estas valoraciones inmediatamente adoptadas para mejorar el programa.

8. A pesar de la incertidumbre sobre la implementación del programa, la misión de ET es muy ambiciosa. El propósito

del programa no es solamente incrementar las ganancias (y la renta) de los beneficiarios de ayudas públicas o reducir el gasto público. El objetivo es convertir a los receptores de ayudas en ciudadanos económicamente independientes y socialmente productivos.

9. Finalmente, el liderazgo político del Estado entiende que las prestaciones de ayuda social son demasiado bajas y en consecuencia ha decidido elevar los niveles de las prestaciones y ampliar los criterios de elegibilidad.

Un investigador debería tener el cuidado, en el momento de diseñar un experimento para evaluar ET, de no introducir nada que pudiera cambiar ninguna de estas características. Cualquier programa de bienestar, formación y empleo con estas características podría producir resultados que serían susceptibles de muchas explicaciones alternativas. ¿Cómo podría pues evaluarse? Más concretamente, ¿cómo se podría diseñar un experimento controlado que eliminara tantas de estas explicaciones alternativas como fuera posible con el objeto de determinar el impacto neto de ET?

La creación de un grupo de control

La primera tarea sería el desarrollo de un mecanismo para asignar aleatoriamente participantes a grupos de tratados y de control. Al fin y al cabo, comparar las tasas de empleo de participantes en ET (voluntarios) con no participantes (no voluntarios) no elimina una explicación obvia de las diferencias: que los participantes en el programa pueden tener mayor éxito simplemente porque estaban más motivados. Para juzgar el impacto neto de este programa sólo - y eliminar tantas explicaciones alternativas como fuera posible - sería útil asignar aleatoriamente a algunos de los voluntarios al grupo de tratados y el resto al grupo de control.

Pero estos voluntarios son seleccionados a través de una agresiva campaña de *marketing*. Una parte se hace indiscriminadamente con panfletos y videos. La mayor parte se hace persona a persona con un trabajador social que explica a cada posible participante los beneficios del programa, tratando de convencer a esta persona que ET puede mejorar su vida de forma significativa. Pero si esta persona participa en un experimento controlado, el discurso del trabajador social estaría sujeto a algunas limitaciones:

“Este programa es realmente super. Deberías apuntarte. Te ayudará a vivir una vida mejor. Tus ingresos subirán. Estarás orgullosa de tí misma. Tus hijos te respetarán más. Debes apuntarte ... Luego, los que llevan el

programa, tiran una moneda al aire, y si sale cara, entras.”

Una de las cosas que el DB aprendió sobre *marketing* fue que tan pronto como un trabajador social conseguía interesar a su cliente en ET, debía apuntarlo sin dilación. No debía enviarlo a casa y esperar que llamara a la oficina de ET para pedir una cita sino que debía acompañar a esta persona personalmente a la oficina de al lado para presentarla a los trabajadores sociales de ET. Ello sugiere que la mera existencia de un experimento controlado puede afectar negativamente a uno de los componentes claves de ET; hasta el voluntario que obtiene cara y es admitido en el programa puede ver su motivación afectada por la incertidumbre, sobre si finalmente obtendrá la ayuda del DB. Operativamente, la asignación aleatoria es incompatible con un programa voluntario que depende de una forma considerable del *marketing* directo, persona a persona, trabajador social-beneficiario, seguido por la entrada rápida en algún componente del programa.

Además, hay un problema ético en la negación de servicios a una persona. Si el Estado no tiene suficientes recursos para proporcionar servicios de formación y empleo a cada beneficiario, la asignación aleatoria puede ser una forma de determinar quien los obtiene. Pero si el Estado tiene suficientes recursos, ¿cómo puede negarlos a alguien que quiere un empleo? Tomando el argumento ético, Richard P. NATHAN, del Rockefeller College, State University of New York, y presidente de Manpower Demonstration Research Corporation (MDRC), escribe: “si una política es universal... difícilmente se puede argumentar que debería suspenderse para algunos participantes que tienen derecho a ella en base al interés científico”¹². Este dilema ético tuvo su influencia en el diseño de la evaluación por MDRC del programa WORK en Arkansas. MDRC utilizó asignación aleatoria “en dos de los ocho condados iniciales, Jefferson y Pulaski South [Little Rock], las únicas áreas en las que el número de personas interesadas en el programa superaba de tal forma las plazas disponibles, que era posible utilizar esta técnica”¹³. Para un programa de ayuda a madres menores de edad, MDRC utilizó una aproximación cuasi-experimental porque la asignación aleatoria, como observa NATHAN, “no era factible por razones políticas”¹⁴.

Por supuesto, los líderes políticos de ET podrían vender el programa con una campaña de *marketing* agresiva y al mismo tiempo negar la participación en el programa a la mitad de los voluntarios. Y los trabajadores sociales podrían mantener un discurso personalizado y entusiasta frente a cada cliente sabiendo que sólo tiene un cincuenta por ciento de posibilidades de entrar en el programa. Pero, incluso en este caso, los asignados al grupo de control pueden recibir servicios similares a los del programa ET.

De hecho, cualquier mujer asignada a un grupo de control tiene acceso a servicios casi idénticos. Sólo necesita ser perseve-

rante en su búsqueda. Si de verdad quiere ayuda pública para encontrar un trabajo, puede acudir a las oficinas de empleo. O, si quiere formación, puede buscar una organización que la proporcione con financiación federal de la *Job Training Partnership Act*. Cada una de las agencias JTPA que tiene además un contrato con ET tiene la obligación de servir (además de a los beneficiarios del grupo de tratados) otros beneficiarios de asistencia social para cumplir la cláusula contractual por la que deben mantener el mismo nivel de esfuerzo. Así pues, estas agencias también seleccionan a beneficiarios del DB y por tanto les resulta relativamente fácil seleccionar a personas que ya han manifestado su interés por los programas del DB.

Por último, consideremos al trabajador social que acaba de intentar convencer a un receptor de ayudas sociales sobre los beneficios y alegrías que le puede proporcionar la formación y el empleo en el momento de descubrir que la moneda lanzada al aire acaba de asignar a esta mujer al grupo de control. La decepción afectará, no sólo al beneficiario sino también al trabajador social. Este (o el director del programa) se enfrenta a lo que NATHAN llama "el dilema del director del programa". ¿Debe el trabajador social decir al receptor de ayudas sociales donde puede encontrar la oficina más cercana de JTPA?¹⁵ ¿Cuántos trabajadores sociales resistirán la tentación?

La definición y la estandarización del tratamiento

Además de la dificultad de crear un grupo de control, también existe el problema de definir el tratamiento y de garantizar su uniformidad. ¿Son tratamiento las actividades de orientación, educación básica, formación profesional y colocación, o también deben incluirse los cambios en las actitudes de los beneficiarios, trabajadores sociales y empresarios? Si el único tratamiento que cuenta es la prestación directa de un servicio social específico, la tarea de evaluación se simplifica considerablemente.

En cualquier caso, ¿cómo puede saberse si la calidad del tratamiento es uniforme? Por ejemplo, si el grupo de tratados recibe servicios de formación, ¿en qué medida son consistentes y lo bastante intensos como para garantizar que todos aquellos que son tratados son realmente tratados y reciben los mismos servicios?

En el caso de la vacuna contra la polio o del impuesto negativo sobre la renta, el tratamiento es concreto y relativamente sencillo de definir y controlar. Es una acción simple: inyectar una vacuna en el brazo de un individuo; inyectar dinero en la cuenta bancaria de una familia. La calidad de la vacuna contra

la polio puede controlarse y mantenerse constante. Si una familia dependiente de las ayudas sociales recibe billetes nuevos, o viejos y arrugados, o un cheque, ello no afectará su comportamiento económico. Por lo tanto, llevar a cabo un experimento controlado para vacunas y políticas sociales directas (como dar dinero a una familia) es *relativamente* fácil. Esta facilidad es sólo relativa. Mantener estándares uniformes para los pagos de AFDC es un problema considerable, y los esfuerzos por controlar la calidad y mantener baja la tasa de errores absorben un número considerable de recursos en todos los departamentos de bienestar. Aún más, los muchos volúmenes escritos sobre los experimentos con impuestos negativos revelan que la evaluación de esta comparativamente simple política social no fue tan fácil¹⁶.

Un programa de bienestar que quiere poner la gente a trabajar puede ser relativamente sencillo. Puede consistir en una determinada actividad relativamente simple —ej., búsqueda de trabajo en grupo—. Incluso para un programa tan modesto, sin embargo, la definición y estandarización del tratamiento puede ser bastante compleja. Por ejemplo, la calidad de las actividades de búsqueda de empleo podría variar considerablemente. Los mismos técnicos en ocupación podrían, en unos casos, tener una clara influencia mientras en otros apenas tener impacto alguno.

¿Cómo podría saberse? Los investigadores podrían comparar distintos trabajadores sociales y distintas actividades de búsqueda de empleo. Pero esto generaría un experimento dentro del experimento. Si algunas actividades producían resultados y otras no, los evaluadores podrían determinar si estas diferencias son estadísticamente significativas. Pero, entonces ¿qué?, ¿deberían los responsables de la política ignorar los resultados del tratamiento ineficaz?

Esta aproximación puede tener sentido para diferentes actividades de búsqueda de empleo. El experimento revelaría que algunas actividades funcionaron y otras no. Pero, ¿deberían tenerse en cuenta los resultados en lo relativo a los trabajadores sociales ineficaces? Esto parece menos legítimo. Puede ser aceptable si las características de los trabajadores sociales ineficaces pueden (a) ser claramente identificadas, y (b) eliminadas en el futuro, ya sea a través de la selección de nuevos trabajadores sociales o de su formación. De lo contrario, los resultados sólo serían aplicables a los mejores trabajadores sociales y no a la media. No serían más legítimos de lo que lo son los de los "investigadores" de la parapsicología que no tienen en cuenta los resultados de sus experimentos cuando consideran que el sujeto no tiene "talento", o en el caso de que lo hubiera tenido anteriormente, tuvo un "mal día".

Además, nada de lo afirmado hasta ahora garantiza que las actividades de búsqueda de empleo en un lugar funcionen en otro. Es muy posible que algunos aspectos de una determinada

actividad sean adecuados para la economía de San Diego en 1982-1984, mientras que otros —que no tuvieron impacto— funcionen bien en New Jersey en 1984 -1986. ¿Pueden pues los resultados de San Diego 1982-1984 aplicarse en Arkansas en 1990 o en San Diego en 2005?

Todos estos problemas de estandarización del tratamiento ponen de relieve por qué un 6% de diferencial en el nivel de ocupación de los participantes en el programa de San Diego es significativo¹⁷. En San Diego se dieron forzosamente varios tratamientos distintos. Durante los dos años y tres meses que duró el experimento, tuvo que haber diferencias significativas en la calidad —y hasta en el carácter de los servicios de búsqueda de empleo y experiencia de trabajo que tuvieron los participantes—. El programa y su evaluación empezaron al mismo tiempo; no era por tanto posible solucionar los fallos inevitables o garantizar uniformidad en los servicios. Dado que descubrir un 6 por ciento de diferencia requería una prueba positiva de que se debía rechazar la hipótesis nula de que no hubo impacto, dado que algunos servicios necesariamente tuvieron poco impacto en la capacidad de los beneficiarios para encontrar un empleo, y dado que muchos en el grupo de tratados recibieron pocos servicios, que MDRC pudiera detectar algún impacto positivo fue realmente significativo.

El efecto Rosenthal

Comparado con el programa de San Diego, ET CHOICES era bastante complejo dado que proporcionaba una amplia variedad de servicios de orientación, educación básica, formación y colocación. Además, estos servicios eran distribuidos por un número considerable de oficinas de bienestar, oficinas de empleo y pequeñas empresas contratadas por el Estado. Algunas de estas oficinas y agentes trabajaban bien. Otros no. La uniformidad es difícil de conseguir —salvo en el caso de que se establezcan controles burocráticos estrechos pero que presionan la uniformidad hacia el común denominador más bajo en cualquier escala de eficacia—.

Además, ET también intentó influir en las actitudes y comportamiento de los receptores de asistencia social *así como* en el comportamiento de cualquier persona que entrara en contacto con éstos —trabajadores sociales, educadores y empresarios—. Por ejemplo, cuando el vicepresidente de una empresa manufacturera dijo a la prensa: “Esta gente realmente quiere trabajar... tenemos a cuatro ganadores aquí, y nos gustaría tener más como ellos”¹⁸, realmente ayudó a romper el escepticismo normal de otros empresarios en torno a contratar personas dependientes de las ayudas sociales.

Si los periodistas claramente informaran que “esta gente”, “ganadores”, y “más como ellos” eran los receptores de ayudas

sociales que habían completado el programa ET, esta publicidad y cualquier cambio de actitudes resultante no habría complicado un experimento controlado. Pero si otros empresarios interpretaban que cualquier persona dependiente de las ayudas públicas podía ser uno de estos “ganadores”, entonces se produciría un problema. Este componente del “tratamiento” —influir en la actitud de potenciales empleadores— podría ayudar a encontrar un empleo, no sólo a los tratados sino también a los controles. Un receptor de AFDC podría sentirse motivado por la campaña agresiva de *marketing* de ET, ser asignado a un grupo de control, luego irse a la oficina de empleo de la esquina, transferida a una agencia de colocación (que tiene que mantener unos determinados niveles de esfuerzo), y finalmente conseguir un trabajo —porque el empresario ya no piensa que es arriesgado contratar a una persona que ha estado recibiendo ayudas para alcanzar un nivel mínimo de ingresos—.

¿Cuántas veces se podría dar esta situación? No muy a menudo. Sólo sería una pequeña dificultad para poder llevar a cabo un experimento controlado de ET CHOICES. Sin embargo, si el diferencial entre el grupo de tratados y el grupo de control es solamente de 6 puntos, varios pequeños problemas como éste, conjuntamente, sí podrían convertirse en un problema importante¹⁹.

Además, el componente del programa que intenta influir en las actitudes puede tener un impacto potencialmente mayor. La actitud con la que se prestan los servicios puede tener tantas consecuencias como los mismos servicios. Los trabajadores sociales, ¿son amables, serviciales y abiertos? ¿O son desagradables, cerrados y despreciativos? La actitud del trabajador social puede enviar señales inequívocas al beneficiario. “Puedes hacerlo. Puedes encontrar un empleo”, o, “No sirves para nada. Nadie te va a contratar.” Uno de los componentes de ET que puede haber sido más importante es el esfuerzo realizado para influir en las actitudes de los empleados del DB.

Concretamente, los directivos del DB intentaron convencer a sus trabajadores sociales sobre la importancia de las siguientes ideas claves:

I. Nuestra misión es ayudar a los receptores de ayudas públicas a encontrar un empleo y a ser económicamente independientes.

II. La mayoría de los receptores de ayudas pueden —si reciben la educación básica y formación necesarias— encontrar y mantener un trabajo.

III. Para ello, es necesario tratar a cada beneficiario como a un individuo y diseñar un programa individualizado que atienda sus necesidades.

Si tuviera que hacerse un experimento controlado, y la actitud es parte del tratamiento, tendría que añadirse un nuevo elemento a este mensaje.

IV. Por supuesto, sólo debéis tener esta actitud cuando habléis con beneficiarios que han sido asignados al grupo de tratados. Cuando tratéis con los del grupo de control, ser los mismos de antes, indiferentes.

Obviamente, es imposible retener el componente actitudinal del grupo de tratados y evitar su aplicación al grupo de control. Si la actitud con la que se presta el servicio es importante, y si los gestores de ET quieren influir sobre las actitudes, necesitan inyectar esta nueva actitud a todas las comunicaciones con los trabajadores sociales del DB.

La actitud con la que se presta un servicio puede ser crítica. Hace un cuarto de siglo Robert Rosenthal de la Universidad de Harvard descubrió que las expectativas de un maestro pueden influir en lo que aprende un niño. Rosenthal administró un test a una serie de niños, asignó aleatoriamente el 20% de cada clase al grupo de tratados, y después dijo a sus maestros que los resultados del grupo de tratados “en el test de expansión intelectual²⁰ indicaban que mostrarían un avance intelectual por encima de lo normal durante el siguiente año académico”. Ocho meses después, Rosenthal volvió a administrar el test: los niños a cuyos maestros se les dijo que tendrían un avance intelectual superior, efectivamente, habían avanzado más. El tratamiento directo fue lo que Rosenthal dijo a los maestros. Pero también hubo un tratamiento indirecto: como los maestros trataron a los alumnos en el grupo de tratados comparado con como los mismos maestros trataron a los controles²¹.

De igual forma, los directivos del DB intentaron aplicar ambos, un tratamiento directo (lo que los gestores dijeron a los trabajadores sociales) y un tratamiento indirecto (como los trabajadores sociales trataron a los beneficiarios). El tratamiento directo eran los mensajes con la intención de convencer a los trabajadores sociales en el sentido de que podían ayudar a sus clientes a encontrar empleos y que éstos podían perfectamente ser productivos en estos trabajos. A su vez, estos trabajadores sociales aplicarían el tratamiento indirecto al trabajar con los beneficiarios de las ayudas.

El efecto Hawthorne

El efecto Rosenthal en educación, al igual que el efecto placebo en medicina, tienen un equivalente en gestión: el efecto Hawthorne. En 1924, en un proyecto para determinar como la iluminación influye en la producción industrial, la Academia Nacional de Ciencias y la Western Electric llevaron a cabo una serie de experimentos en la planta de Hawthorne, una fábrica de Western Electric en los alrededores de Chicago. Los directivos iluminaron una parte de la fábrica. El grupo de tratados incrementó su productividad. Pero lo mismo hizo el grupo de control. Otra vez, se incrementó la iluminación del grupo de

tratados. Otra vez, aumentaría la productividad. Nuevamente para los dos grupos. Luego, los directivos apagaron las luces. Todavía la productividad aumentaba - y para los dos grupos. Obviamente, los cambios en la iluminación no influían en la productividad. Más bien, lo que parecía marcar las diferencias era la atención que los directivos daban a ambos grupos — tratados y controles—. Para llevar a cabo el experimento y recoger datos, los directivos tuvieron que dedicar mucha atención a los participantes en ambas secciones de la fábrica. Esta atención fue el factor que marcó la diferencia entre lo que pasó antes y después del experimento²².

La investigación médica intenta distinguir entre los tratamientos que tienen un efecto fisiológico y los que tienen un efecto puramente psicológico. ¿Por qué pagar por un medicamento caro cuando una pastilla de azúcar tendrá el mismo efecto? El deseo de eliminar el efecto placebo cuando se prueban nuevos medicamentos no significa que se deba eliminar de la práctica de la medicina. Ciertamente, el efecto placebo es un elemento importante entre los instrumentos de los que dispone un médico. “Toma esta medicina. Te curará.” De la misma manera que el efecto Rosenthal es una herramienta importante para el maestro: “Puedes aprenderlo. Sé que puedes”. Médicos y maestros no sólo necesitan aplicar las últimas tecnologías de sus profesiones: también tienen que influir en la actitud del individuo al que quieren curar o enseñar.

Lo mismo se puede decir de las políticas públicas. No todas las herramientas son tecnológicas. Algunas de las más importantes pueden ser psicológicas. Para que los trabajadores sociales puedan ayudar a los beneficiarios de ayudas públicas a ser financieramente independientes, parte de su trabajo es técnico: enseñar al beneficiario a trabajar con una máquina, por ejemplo. Pero otra parte de la tarea es psicológica: convencer a aquella persona de que puede aprender a utilizar aquella máquina, puede entenderse con el empresario, conseguir poner a los niños en una guardería y solucionar las inevitables crisis personales que surgirán, y aun llegar al trabajo todos los días puntualmente.

Los responsables de los departamentos de bienestar que intentan convencer a los empresarios para que contraten a estas mujeres también tienen una tarea a la vez técnica y psicológica. Tienen que demostrar concretamente que los participantes en sus programas tienen las habilidades necesarias. También tienen que convencer a estos mismos empresarios que no tendrán problemas para trabajar con estas personas hasta ahora dependientes de las ayudas sociales. De hecho, parte del trabajo de los gestores públicos consiste en hacer que estos empresarios se sientan satisfechos de contratar personas procedentes del DB. La parte psicológica del tratamiento puede ser más significativa que la parte tecnológica. El directivo de cualquier programa de formación y empleo quiere sacar el máximo provecho del efecto Rosenthal.

En el mismo sentido, cualquier responsable de un programa de bienestar (de hecho, cualquier gestor público) quiere sacar el máximo provecho del efecto Hawthorne. Si prestar atención a los trabajadores sociales tiene un impacto en su eficacia, los gestores deben hacerlo. Si salir en los medios de comunicación hace que aquellos que trabajan en el programa, sus beneficiarios, y aquellos que podrían contratarles, sientan que la sociedad está prestando atención a lo que hacen, entonces los gestores deben hacer todo esto. Quizás, la imposición de la obligación de buscar trabajo, o la formación, o cualquier otra característica de un programa de formación y empleo no es tan importante como la iluminación en la planta de Hawthorne. Quizás lo que cuenta es la atención que los gestores prestan a estos individuos cuyas acciones son necesarias para que el programa funcione.

Si el tratamiento puede ser administrado individualmente e impersonalmente a cada miembro del grupo de tratados - y puede ser al mismo tiempo positivamente negado a los miembros del grupo de control - entonces tiene sentido llamar a la diferencia entre los resultados para los grupos de tratados y de control "impacto neto." Pero si el tratamiento no puede ser negado al grupo de control, es difícil argumentar que no encontrar ninguna diferencia entre tratados y grupo de control - el no poder refutar la hipótesis nula - también significa que el programa no tuvo impacto alguno. Además, si para conseguir los objetivos del programa sus gestores lo han diseñado de tal forma que facilita el acceso múltiple a los tratamientos, es absurdo modificar significativamente algunos de sus componentes sólo para facilitar la evaluación de otros.

¿Es que las políticas sociales más eficaces sólo constan de un elemento crítico? Casi nunca. Atacar problemas como la dependencia económica del Estado requiere políticas con múltiples componentes (y con la flexibilidad de poder tratar individuos distintos con componentes distintos). Además, varios de estos componentes pueden ser necesarios para poder demostrar que la política globalmente considerada es efectiva. Para estas políticas, es más difícil determinar impacto neto. Y nuevamente, sería un error limitar nuestras opciones de políticas sociales a aquellas que las ciencias sociales pueden evaluar con mayor facilidad.

Ser el primero de la cola

Durgin Park es uno de los restaurantes más conocidos de Boston. Situado detrás de Faneuil Hall, ésta fue durante muchos años una zona deprimida de la ciudad. Pero cuando la zona de Quincy Market fue renovada, el local mejoró significativamente. La fórmula, sin embargo, no cambió: unas gruesas rebanadas de *roast beef*, precios moderados, largas mesas con manteles de

cuadros rojos, camareras poco amables y largas colas. Pero había un secreto —todo el mundo lo sabía—. Si comprabas una bebida en el bar, te ponían delante de la cola del comedor.

Este sistema no permite servir más comidas, a más personas, o más rápidamente. Simplemente, cambia la cola. La gente que espera en la cola tiene que esperar más. El comedor tiene un número fijo de sillas. Quizás esto es todo lo que hace cualquier programa de bienestar, formación y empleo. Quizás en San Diego algunos de los beneficiarios en el grupo de control que no obtuvieron empleos hubieran encontrado trabajo sin el programa. Quizás lo que hizo el programa fue poner a los participantes delante de la cola, mientras se quedaban sin empleo los del grupo de control.

Supongamos que un experimento controlado de ET CHOICES muestra un impacto neto. Todavía se pueden dar dos explicaciones alternativas: (1) los participantes salieron del programa con mejores habilidades y mejores actitudes hacia el trabajo, o (2) el *marketing* y los servicios de colocación del programa convencieron a potenciales empleadores sobre la oportunidad de contratar participantes en ET antes que otros demandantes de empleo. Así pues, los participantes en ET habrían desplazado a otros beneficiarios del DB de estos empleos. O pueden haber desplazado a mujeres casadas que ahora no podrían encontrar un empleo con el que completar los ingresos de sus familias.

Si el número de personas que quieren y pueden realizar los trabajos más sencillos excede la oferta para estos empleos de manera que (después de un corto período de transición) cada uno de estos empleos es tomado, luego, cualquier programa de formación y empleo puede conseguir poco. Influye en *quienes* consiguen estos empleos, pero no contribuye en nada a la economía (a no ser que mejore *como* estos trabajadores desempeñan su trabajo). El programa puede, sin embargo, reducir los gastos en subvenciones individuales y afectar la distribución de la renta. La mujer no casada que consigue el empleo a través del programa dejaría la asistencia pública, incrementando su renta y eliminando su caso del presupuesto; puede conseguir un trabajo que de otra forma sería para mujeres casadas que tendrán que quedarse como amas de casa en familias con un nivel más bajo de renta (un solo salario) pero que, sin embargo, está por encima del nivel que permite recibir ayudas públicas.

Si, en cambio, la oferta de personas que quieren y pueden realizar los trabajos más sencillos excede la demanda, pero no se consiguen las personas necesarias para estos empleos (porque los empresarios no pueden encontrar a los potenciales empleados o al revés), entonces un programa de formación y empleo podría ser bastante importante. La actividad de colocación de estos programas puede proporcionar a los empresarios trabajadores que de otra forma no habrían encontrado, de la misma manera que puede proporcionar trabajo a personas que de otra forma estarían dependiendo de las ayudas sociales. O si

muchos de los puestos de trabajo no se llenan por la falta de un número suficiente de demandantes con un nivel mínimo de cualificaciones, la formación de los beneficiarios tendrá algún tipo de impacto positivo en el empleo, el presupuesto social y la renta familiar en el Estado. De hecho, tendría un efecto positivo al contribuir al crecimiento de la economía del Estado. (En cualquier caso, el "Milagro Económico" de Massachusetts, del que hablaba a menudo el Gobernador M. Dukakis durante la campaña presidencial, dependía menos de los participantes en programas de formación como ET que de los graduados del MIT).

Un programa educativo —ya sea una empresa de formación con financiación de JTPA o un programa universitario de administración de empresas— puede contribuir a encontrar trabajos para sus graduados, pero ello no significa que consiga un objetivo social. La formación puede proporcionar conocimientos y habilidades que permiten desempeñar mejor un trabajo. O la formación puede que haga poco más que dar una credencial que sirve a los graduados de un programa para ponerse delante de la cola de los que buscan empleo. Una vez en el trabajo, los candidatos con "más formación" puede que no lo hagan mejor que otras personas con experiencias similares pero sin esta formación o credencial.

En cualquier caso, para simplificar su tarea de búsqueda de nuevos empleados, los empresarios pueden decidir (arbitrariamente aunque no sin razón) considerar sólo a aquellos que tienen la credencial. Porque éstos funcionan adecuadamente en el trabajo (debido a sus antecedentes, no su formación) los empresarios pueden limitar la selección entre los que tienen la credencial. Ello contribuirá a aumentar la demanda por la credencial de formación (y por tanto su precio). Al final, se ha convertido en la entrada que hay que pagar para conseguir trabajo (igual que comprar una bebida en el Durgin-Park es la entrada que hay que pagar para entrar en el comedor). Pero, no contribuye en mucho a la productividad económica de la sociedad globalmente considerada.

Incluso si un experimento controlado para ET revelara un impacto neto, todavía puede darse una explicación alternativa. ¿Es el programa poco más que un Durgin Park formativo que invita a una bebida cara en el bar a los que tienen los antecedentes, habilidades y actitudes necesarias, antes de ponerlos delante de la cola del empleo? Ningún experimento controlado puede responder a esta pregunta²³.

El concepto del experimento social es muy válido y sin embargo a la vez simple: aleatoriamente divide a los participantes en dos grupos. Dale a un grupo el tratamiento y no permitas que el otro lo reciba. Luego utiliza los tests estadísticos para ver si hay alguna diferencia real en el nivel de éxito.

Si el tratamiento es un medicamento, el experimento puede ser muy útil. El tratamiento es un acto sencillo y simple que

puede darse o retirarse con total seguridad y la fórmula no cambia. Además, el número de éxitos o fracasos no es fijo. Si un individuo no contrae una enfermedad, la probabilidad que otro la contraiga no aumenta (si bien en el caso de una epidemia, las probabilidades de que enfermen los dos grupos aumenta).

Pero si el tratamiento es un programa social con múltiples componentes, un experimento puede que no sea tan eficaz para distinguir entre causa y efecto. Por ejemplo, puede que sea imposible que los controles no reciban algunos componentes clave del programa. En esta situación, los resultados del experimento pueden subvalorar el impacto de un programa. O, si el éxito buscado es un trabajo a tiempo completo con un salario decente y seguridad social, el número de éxitos puede que sea fijo, por lo que todo lo que hace el programa es influir en quien consigue estos éxitos. En esta situación, los resultados de un experimento pueden sobrevalorar el impacto de un programa. Además, cuando los líderes políticos que crearon el programa aprenden algo que puede mejorarlo, rápidamente introducirán los cambios necesarios.

Los responsables políticos a los que les agrada sonar la musiquilla del Grupo de Control deberían saltarse el primer baile y pensarse seriamente lo que puede o no puede probar un experimento social en el caso concreto de un programa específico con características específicas.

¿Cómo y cuándo medir el éxito?

Poner a sus participantes delante de la cola del empleo no es suficiente para ET. La misión del programa no era aumentar los ingresos de los participantes o reducir el presupuesto en ayudas sociales. Los creadores de este programa tenían unos estándares más elevados. Su misión era ayudar a sus clientes a escapar de la pobreza para siempre. En consecuencia, sería inapropiado evaluar este programa sobre la base de comparar la media de ingresos de los tratados y de los controles. Un grupo de tratados podría tener estadísticamente ingresos más elevados y sin embargo estar menos cerca de alcanzar la independencia económica.

Para evaluar ET CHOICES en términos de la misión adoptada por sus creadores, un evaluador debe comparar los grupos de tratados y de control en términos de su estado de bienestar. ¿Cuántos aún reciben ayudas? ¿Cuántos han pasado a ser financieramente independientes? El evaluador necesita mirar estas diferencias, no sólo a los treinta días o un año sino también, pongamos, cinco años.

Esto plantea la cuestión que Lester SALOMON, de John Hopkins University, llama "la dimensión temporal de la evaluación". ¿Cuánto debería evaluarse este programa? ¿Puede un evaluador po-

ner a prueba los resultados de un programa inmediatamente? O debe esperar los efectos retardados —como el retorno a la dependencia del Estado de aquellos que, pongamos, trabajaron dos años—. Ciertamente, el efecto Rosenthal en las actitudes puede durar muy poco. Al contrario, pueden darse “efectos latentes” —algunos de los que primero tuvieron trabajos a tiempo parcial, luego encontraron empleos a tiempo completo— al igual que efectos “adormecidos”— aquellos que encontraron trabajo dos años después de terminar su formación. ¿Debería el evaluador esperar?, ¿durante cuánto tiempo?²⁴

La evaluación por MDRC del programa Redirección para madres menores de 20 años, ilustra la importancia de la temporalidad en evaluación. Después de un año, las jóvenes que recibieron los servicios del programa estaban en mejor situación que el grupo de control, pero “esta ventaja claramente desapareció al cabo de dos años”. En una evaluación retrospectiva a los cinco años, sin embargo, MDRC encontró “algunos resultados importantes”. Las madres que participaron en el proyecto “estaban trabajando más horas por semana y tenían ingresos más elevados; eran menos propensas a recibir ayudas públicas; y sacaron mejores resultados en un conocido test sobre habilidades parentales”. Además los niños en el grupo de tratados “mostraban mejores habilidades cognitivas, medidas en un test de conocimiento de vocabulario: y tenían menos problemas de comportamiento”²⁵. Los servicios que proporcionaba el proyecto, concluía J. Gueron, presidente de MDRC, pueden ser importantes a largo plazo²⁶.

¿Cuánto tiempo hay que esperar? Concretamente, ¿cuánto debe esperar un evaluador si los gestores del programa quieren mejorarlo?, ¿debería esperar dos años?, ¿es bastante para que los gestores corrijan el programa, o al menos, corrijan los aspectos fundamentales? O, si el programa es muy complicado, ¿se debe esperar cinco años? El programa —cualquier programa— no será nunca perfecto. Sus gestores siempre utilizarán nueva información y nuevas capacidades para mejorarlo. ¿Cuándo es mejor evaluar?

El programa inicial en San Diego de búsqueda de empleo y experiencia real de trabajo terminó igual que empezó. Las principales características del programa no cambiaron. En consecuencia, la evaluación y el programa podían (y de hecho) empezaron al mismo tiempo. Lo que se evaluaba era pues lo que sus creadores realmente querían hacer.

Al contrario, en el caso de ET CHOICES el programa evolucionaba constantemente. Nadie pretendió que lo que se hizo el primer año sería la última versión del programa. El objetivo —como solían señalar sus gestores— era poner en marcha el programa y *después* arreglarlo. Desde el principio, la gestión individualizada de cada caso era un componente esencial de ET. Sin embargo, hasta el quinto año de funcionamiento, no fue una realidad.

Un responsable del programa siempre puede decir, “No, aún no. Acabamos de introducir algunas modificaciones importantes. No se puede evaluar hasta que estos cambios estén bien asentados.” Pero el programa nunca será perfecto. Siempre estará cambiando. E. Fitzgerald, el primero en llamar la atención sobre los sobrecostos del avión de carga C5A para el Ejército del Aire, ha resumido este problema de la siguiente forma: “Sólo hay dos fases de un programa: la primera es ‘demasiado pronto para poder decir nada.’ La segunda: ‘demasiado tarde para pararlo’”²⁷.

El juego de los números

Cualquier directivo público —especialmente el que acaba de incorporarse a un programa— necesita producir resultados. Necesita mostrar al legislativo, a la prensa y a todos los que tienen algún interés en el programa, que puede conseguir lo que se propone. También tiene que mostrar claramente a los que trabajan en su organización que le importan sus objetivos y que juntos pueden conseguir algunos resultados significativos. A menudo, necesita un éxito que, aunque sea menor, le dé visibilidad externa.

La manera más fácil de hacerlo consiste en fijar un objetivo bien visible e intentar conseguirlo. El objetivo debe ser significativo: también debería ser difícil de conseguir. Sin embargo, si el gestor es listo, escogerá un objetivo que parece más difícil de lo que realmente es. En Octubre de 1983, el Comisionado del DB, M. Atkins, se propuso conseguir 6.000 empleos para beneficiarios de AFDC en nueve meses. Si bien, para muchos, esto parecía imposible, Atkins vio que el objetivo era perfectamente alcanzable. Atkins sabía que podía pedir a la Oficina de Seguridad en el Empleo colocaciones “porque (1) Kristin [Demong, su esposa] estaba allí [era la directora] y (2) era más fácil y más rápido colocar a la gente en un empleo que formarles.” En junio de 1984, cuando se comunicó que el objetivo se había conseguido, se envió un mensaje muy potente sobre las intenciones y capacidad de gestión del DB tanto interna como externamente.

Obviamente, éste no es el único estilo por el que puede optar un directivo. Muchos intentarán evitar este “juego de números con luz y taquígrafos”. Intentarán conseguir empleos para los beneficiarios del DB. Pero se resistirán a fijar objetivos específicos para el consumo interno o externo. También se resistirán a anunciar públicamente que han alcanzado los objetivos que se habían fijado privadamente. Ciertamente, cuando un directivo público empieza a jugar el juego de los números, no puede parar. Si ha demostrado su capacidad para encontrar empleo a 6.000 beneficiarios en nueve meses, tendrá que repetir el truco. Siempre tendrá que responder a la cuestión: ¿Qué

has hecho para mi últimamente? (En el sector privado, por supuesto, las opciones son pocas. Tienen que preocuparse por el número crítico —beneficios— cada año.) Algunos gestores públicos preferirían no tener que entrar nunca en este juego.

Además, hay otro riesgo. Sopongamos que el objetivo seleccionado para la primera pequeña victoria no es el adecuado para el segundo o el tercer año. El uso de un solo objetivo puede distorsionar el comportamiento organizativo. También puede distorsionar las expectativas de los ciudadanos. Si se fija la atención en colocaciones directas, nadie prestará atención a las complicaciones que entraña la formación, que, después del primer éxito, puede pasar a ser la parte central del programa. La prensa continuará informando sobre historias que sean simples, personales y simbólicas.

Centrar la atención externa en el número de casos cerrados puede resultar peligroso a largo plazo. En el caso de ET era inevitable que la atención se centrara en el número de personas dependientes de AFDC, incluso si el Gobernador Dukakis no hubiera dado tanta importancia a este indicador. En el mismo sentido, aunque el DB no hubiera relacionado tan claramente ET con el número de casos, el legislativo lo habría hecho. Cualquiera esperaría que los empleos conseguidos por el programa se tradujeran en menos personas dependientes del Estado. La utilización de esta medida en público por el Gobernador, cuando el programa solamente llevaba dos años, hizo que muchos utilizaran el número de casos para evaluar ET. Puesto que ET era sólo uno de los múltiples factores que influían en el número de personas dependientes de las ayudas sociales, y dado que casi todos los demás factores estaban fuera del control del DB, el legislativo o el gobernador, el argumento de que ET era la única causa de que en Massachusetts se redujera el número de personas dependiendo del Estado, tenía que —un día u otro— girarse contra los que lo habían utilizado.

En 1988, J. O'NEILL, del Baruch College, propuso "calcular la medida en que ET había reducido el número de casos y el dinero público gastado por el DB en ayudas sociales". En su análisis, publicado en marzo de 1990, concluía que ET no había conseguido ni lo uno ni lo otro. "Muy poco o nada de la reducción en el número de casos", escribía O'NEILL, "puede razonablemente atribuirse al programa"²⁸. El análisis había sido examinado por algunos expertos. "El análisis de O'NEILL ciertamente muestra como ET no ha reducido el número de casos o el gasto en la medida suficiente para compensar el gasto inmediato que ha supuesto para el contribuyente de Massachusetts," decía Alice RIVLIN, de la Brookings Institution. "Sin embargo", continuaba RIVLIN, "es importante tener en cuenta que hay otras cuestiones como por ejemplo los beneficios a largo plazo para los participantes, que no se tienen en cuenta en este estudio." Efectivamente, el informe de O'NEILL no examinaba estas otras cuestiones. ET había sido creado para ayudar a las madres que

necesitaban las ayudas sociales a ser económicamente independientes. Pero la cuestión del número de casos en la que se había centrado O'NEILL fácilmente atrajo el interés de la prensa. El primero de mayo de 1990, el titular en el *Boston Globe* decía: La solución del Estado al problema del volumen de personas que dependen de las ayudas del DB parece no funcionar". Aquel que vive del número de casos, muere por el número de casos.

La evaluación del Urban Institute

En 1987, el DB contrató al Urban Institute para realizar una evaluación de ET. En su pliego de condiciones, el DB dejaba bien claro que un experimento controlado con asignación aleatoria no era un diseño de investigación aceptable:

ET CHOICES... se basa en la idea de que es el cliente el que debe optar, lo que facilita que se adopte un rol más activo en la prestación del servicio. Hacer de la opción individual la parte esencial de este programa nos sirve para impulsar la idea de que son los propios clientes los que controlan su vida. Así pues, las metodologías que implican la asignación aleatoria de clientes a grupos de control y a varios grupos que reciben los servicios supondría negar la opción por parte del cliente en la que se fundamenta el programa. El Departamento de Bienestar Público considera que la asignación aleatoria es una metodología diametralmente opuesta a la filosofía básica del Programa ET CHOICES"²⁹.

La decisión de utilizar como diseño de investigación un cuasi-experimento³⁰ en lugar de un experimento significaba para D. NIGHTINGALE, L. BAWDEN, y L. BURBRIDGE, del Urban Institute, que hicieron la investigación, que "la evaluación ya ha sido criticada antes incluso de que empezara"³¹.

En el pliego de condiciones, el DB señalaba que les interesaba menos determinar si el programa funcionaba que aprender como hacerlo funcionar mejor. "A los administradores del programa les interesaba especialmente," según NIGHTINGALE y sus colaboradores, "un análisis de los distintos componentes y entre distintos grupos de participantes"³². Efectivamente, el pliego de condiciones planteaba bastantes temas de gestión:

V. ¿Son algunos componentes de ET más eficaces que otros en ayudar a los clientes a salir de la dependencia del DB y permanecer ajenos a las ayudas públicas?

VI. ¿Qué participantes en cada componente consiguen encontrar trabajos y ser independientes económicamente?

VII. ¿Hasta qué punto los clientes consiguen mantener sus trabajos? ¿Cuántos están todavía ocupados un año después de

encontrar su empleo ET? ¿Cuántos continúan siendo independientes un año después?

VIII. ¿Son algunos tipos de clientes más propensos a guardar sus trabajos y mantenerse económicamente independientes?

IX. ¿Hasta qué punto la provisión de servicios de guardería por el programa ET es un factor que ayuda a los participantes a guardar sus trabajos y mantenerse económicamente independientes?³³

“El principal objetivo de la evaluación”, escribió el equipo del Urban Institute, “es determinar qué componentes del programa ET funcionan mejor para qué tipo de clientes AFDC”³⁴.

En setiembre de 1990, el Urban Institute publicó su informe describiendo a ET como “un éxito considerable”. En particular, el estudio encontró:

X. “La participación en ET redujo la duración mediana del período AFDC en un poco más de cuatro meses, cerca de un 29%”.

XI. “La participación en ET redujo la subvención media mensual en algo más de \$25 al mes, una reducción del 8 por ciento”.

XII. “La participación en ET se estima aumentó el empleo... en un 8.2 por ciento (desde 36.3 por ciento a 44.5 por ciento...)”.

XIII. “El efecto medio estimado en los ingresos de los participantes fue de un aumento de \$390 en seis meses, aproximadamente el 34 por ciento”.

Todos estos datos eran estadísticamente significativos a un nivel del 1 por ciento y los impactos estimados eran “un poco superiores a los encontrados en otros proyectos de reforma en otros estados.” Por ejemplo, el efecto en los ingresos (aumento de \$390 en seis meses) era un impacto superior al encontrado para otros programas, que van desde \$100 hasta \$470 por año, durante el primer año de participación. ET no tuvo, sin embargo, ningún “efecto en la tasa en la que antiguos beneficiarios volvían a depender de las ayudas públicas”³⁵.

El Urban Institute no intentó determinar el impacto neto de ET en el número de casos ni tampoco hizo un análisis coste-beneficio. Pero sí analizó con mayor profundidad los efectos de varios componentes de ET, y sus impactos diferenciales eran bastante predecibles:

XIV. La educación básica aumentaba el tiempo que se pasaba en AFDC y la subvención mensual, mientras reducía empleo e ingresos. Los beneficiarios de AFDC que necesitaban educación básica (como Inglés como segunda lengua) eran los menos susceptibles de encontrar empleo: la decisión de obtener educación es una decisión que supone posponer la búsqueda de empleo hasta haber completado el programa formati-

vo. Pero el Urban Institute encontró que “aquellos que se empleaban después de cursar la educación básica permanecían empleados durante más tiempo que los que simplemente buscaban un empleo”³⁶.

La búsqueda directa de empleo tenía “impactos positivos” que “eran mayores que los encontrados en otros estudios” pero de corta duración”. Los participantes en ET que optaban por la búsqueda de empleo “tenían períodos más cortos de empleo que otros participantes en ET”³⁷.

XVI. Los componentes más intensivos ofrecidos por ET —experiencia de trabajo y formación ocupacional— tenían “el mayor impacto observado”. Por ejemplo, trabajo real [apoyado por el DBI] aumentaba los ingresos en seis meses en \$885 y la ocupación en un 42 por ciento³⁸.

“Estos resultados son consistentes con otros estudios”, concluía el Urban Institute, “confirmando que la inversión en actividades más intensivas (y también más caras) vale la pena dado el mayor impacto observado. Así pues, servicios intensivos como la experiencia de trabajo real son bastante caros, pero también sus impactos son mayores”³⁹.

De los trabajos encontrados por participantes en ET, el 48 por ciento incluía seguro sanitario. Además, los trabajos con seguro de sanidad “tenían un nivel de terminación de la relación laboral en torno a un 50 por ciento menor” que los que no tenían seguro—lo que sugería “una diferencia en la duración media del empleo de un 106 por ciento”—. Pero el Urban Institute advierte, correlación no es causación: el seguro sanitario puede que no sea la causa por la que ex participantes en ET dejen estos trabajos menos deprisa. Más bien, seguro sanitario debe de estar correlacionado con ‘buenos’ trabajos”. Puede que los trabajos con seguro sanitario son mejores trabajos; pagan mejor y tienen mejores condiciones de trabajo, estabilidad, y otros beneficios⁴⁰.

Para el evaluador que le preocupan las explicaciones alternativas, la causación en esta correlación es sospechosa. Pero para el gestor público, cuya preocupación es producir resultados, esta correlación tiene en sí misma bastante valor. Para el gestor público, no importa que el seguro sanitario no sea la causa de un empleo por más tiempo. Lo que importa es que el seguro sanitario parece un buen indicador de un buen trabajo. ¿Qué es un “buen trabajo”? ¿Cuáles son las características de un buen trabajo y cómo pueden combinarse en un sencillo y simple índice de calidad del empleo? Puede que estas preguntas requieran otro estudio. Pero el gestor público no necesita esperar. Ya tiene un indicador útil de calidad del empleo: si el empleo incluye seguro sanitario. Y ¿cómo se define un “buen trabajo” en este caso? Para el gestor público la respuesta es bien simple: un buen trabajo es aquel que dura aproximadamente el 106 por ciento más que otros.

Gestores públicos y evaluadores perciben hasta el dato más pequeño de maneras distintas. Esto es así porque tienen distintas responsabilidades, distintas perspectivas y distintas prioridades. Ven el trabajo de diseñar, evaluar y gestionar programas públicos de formas bastante diferentes. Los evaluadores querían diseñar el programa de forma que se pueda evaluar. Pero ET no puede evaluarse —al menos no con la herramienta definitiva del

evaluador, el experimento social— sin cambiar algunas de sus características que sus responsables políticos consideran centrales para sus objetivos y eficacia. No podemos determinar de forma definitiva en qué medida ET ayuda a los beneficiarios de ayudas sociales a salir de su situación de pobreza, más de lo que podemos determinar de forma definitiva la medida en que el bombardero Stealth sirve para disuadir una agresión externa.

Notas

Artículo traducido por Xavier BALLART.

* Duke University.

¹ Traducción del capítulo IX del libro del mismo autor *Leadership Counts*, con el permiso de Harvard University Press. Por razones de espacio, no se ha traducido la totalidad del capítulo.

² N.T. AFDC se refiere al programa de ayuda federal a las familias con dependientes que no llegan a un nivel mínimo de ingresos: «Aid to Families with Dependent Children». Los beneficiarios de este programa suelen ser mujeres que viven solas con niños.

³ Office of Research, Planning and Evaluation, *An Evaluation of Massachusetts Employment and Training Choices Program: Interim Findings on Participation and Outcomes, FY84-FY85* (Boston: Massachusetts Department of Public Welfare, enero 1986, pp.i y 2.

⁴ *Ibid.*, pp. i,ii y 3.

⁵ *Ibid.*, pp. 63-64

⁶ Department of Public Welfare, *The Employment and Training Choices Program: Program Plan and Budget Request, FY 87* (Boston: Commonwealth of Massachusetts, enero 1986), pp. 6-7.

⁷ *Ibid.*, pp.7 y 17.

⁸ Department of Public Welfare, *The Employment and Training Choices Program: Program Plan and Budget Request, FY88* (Boston: Commonwealth of Massachusetts, febrero 1987), p.8.

⁹ Department of Public Welfare, *Program Plan and Budget Request, FY87*, p 16.

¹⁰ N. T. Massachusetts Taxpayers Foundation, en inglés.

¹¹ Massachusetts Taxpayers Foundation, *Training People to Live Without Welfare*, (Boston, Massachusetts Taxpayers Foundation, agosto 1987), pp. ii-iii.

¹² Richard P. NATHAN, *Social Science in Government: Uses and Misuses* (New York: Basic Books, 1988), p. 125.

¹³ Daniel FRIEDLANDER, Greory HOERZ, Janet QUINT, y James RICCIO, *Arkansas: Final Report on the WORK Program in Two Counties* (New York: Manpower Demonstration Research Corporation, setiembre 1985), p. 9.

¹⁴ NATHAN, *Social Science*, p. 80.

¹⁵ *Ibid.*, p. 89.

¹⁶ Ver el capítulo 8, nota 7, del libro de R.D. BEHN, *Leadership Counts*.(Cambridge, Harvard University Press, 1991).

¹⁷ N. T. El autor se refiere al experimento de San Diego, donde el 61% en colocaciones para el grupo de tratados superó en seis puntos el nivel de colocaciones para el grupo de control que fue del 55%.

¹⁸ Katheleen DURAND, «Dukakis ensalza el programa de empleo» *Fall River Herald*, 11 de setiembre, 1985.

¹⁹ NATHAN argumenta en el mismo sentido. La evaluación de un programa piloto de rentas mínimas como los experimentos con impuestos negativos sobre la renta no es tan útil como la evaluación de un programa piloto de prestación de servicios de bienestar, formación y trabajo. Entiende que un programa nuevo de renta mínima universal sólo se podría aprobar después de un «debate nacional...claramente visible». En consecuencia, «se sabría que las reglas del juego habían cambiado» por la nueva legislación, y por tanto, «los que fueran expuestos a esta nueva política cambiarían su comportamiento de maneras que no se podrían anticipar en el

contexto de una investigación en que tal debate no hubiera tenido lugar.» Una pequeña prueba piloto de un programa de renta garantizada no descubriría los cambios en las actitudes y en el comportamiento porque las personas tratadas no serían expuestas a la publicidad inherente a un debate nacional. Al contrario, los programas de prestación de servicios pueden instituirse de formas «más discretas» de forma que «sean percibidos, simplemente como otro programa». NATHAN concluye: «los programas de prestación de servicios normalmente no tienen la visibilidad y publicidad que podría cambiar las expectativas y el comportamiento de un número considerable de personas.» NATHAN, *Social Science*, pp. 61-62 y 194-195.

²⁰ N. T. «Intellectual blooming» en inglés.

²¹ Robert ROSENTHAL y Leonore JACOBSON, «Teachers' Experiences: Determinants of Pupils' IQ Gains», *Psychological Reports*, 19 (1966), pp. 115-118. ROSENTHAL hizo lo mismo con ratas. Robert ROSENTHAL y Kermit L. FODE, «The Effect of Experimenter Bias on the Performance of the Albino Rat», *Behavioral Science*, 8 (1963), pp. 183-189. Robert ROSENTHAL y Reed LAWSON, «A Longitudinal Study of the Effects of Experimenter Bias on the Operant Learning of Laboratory Rats», *Journal of Psychiatric Research*, 2 (1964), pp. 61-72. Los investigadores dieron a estudiantes de psicología ratas aleatoriamente etiquetadas como «listas» y «tontas» para hacer algunas tareas propias de laboratorio. Las ratas que los estudiantes creían haber sido alimentadas para ser ratas listas hicieron mejor su cometido que aquellas que los estudiantes creían haber sido alimentadas para ser tontas, si bien en realidad las ratas habían salido del mismo lugar. El estudiante que creía que su rata era lista «era más entusiasta, agradable...», y estaba más interesado en el rendimiento de la rata. Prefería a la rata lista y la observaba con mayor interés.» ROSENTHAL y LAWSON, p. 71.

²² Los resultados de la prueba de 1924-1927 sobre iluminación del Experimento de Hawthorne aparecieron en C.E. SNOW, «A Discussion of the Relation of Illumination Intensity to Productive Efficiency», *Technical Engineering News*, noviembre de 1927, pp. 257-282. En realidad, sin embargo, el verdadero «Hawthorne Effect» es el resultado de los subsiguientes estudios de cinco mujeres que ensablaban conexiones para teléfonos. Ver F. J. ROETHLISBERGER y William J. DICKSON, *Management and the Worker*, (Cambridge, Mass.: Harvard University Press, 1939.). Estas interpretaciones de los datos han sido puestas en duda por otros analistas. Ver H. MCLIVAN PARSONS, «What Happened at Hawthorne?» *Science*, 1983 (Marzo 8, 1974), pp. 922-932; Richard HERBERT FRANKE y James D. KAUL, «The Hawthorne Experiments: First Statistical Interpretation», *American Sociological Review*, 43 (octubre 1978), pp. 623-643.

²³ Ralph E. SMITH, *Work Related Programs for Welfare Recipients* (Washington, D.C.: Congressional Budget Office, abril 1987), pp. 11, 33-34 y 60-61.

²⁴ Lester M. SALOMON, «Follow-Ups, letdowns, and Sleepers: the Time Dimension in Policy Evaluation», en Charles O. Jones and Robert D. Thomas (eds.) *Public Policy Making in a Federal System* (Beverly Hills, Ca.: Sage publications, 1976), pp257 y 260-261.

²⁵ Denise F. POLIT, Janet C. QUINT y James A. RICCIO, *The Challenge of Serving Teenage Mothers: Lessons from Project Redirection*, (New York: Manpower Demonstration Research Corporation, octubre 1988), pp. 1 y 3.

²⁶ Spencer RICH, «Welfare Mothers Show Benefits of Training», *The Washington Post*, 7 de octubre, 1988.

²⁷ Molly MOORE, «A. Ernest Fitzgerald: Analyst Who knows the Price of Exposing Cost Overruns», *The Washington Post*, 23 de febrero, 1987.

²⁸ June O'NEILL, *Work and Welfare in Massachusetts: An Evaluation of the ET Program* (Boston, Pioneer Institute for Public Policy Research,1990), p.4

²⁹ Office of Research, Planning and Evaluation, *Request for Proposal for an Evaluation of Employment and Training (ET) Choices* (Boston: Massachusetts Department of Public Welfare, agosto 1986), p. 3.

³⁰ Como grupo de comparación, NIGHTINGALE et al. utilizaron una muestra estratificada de no participantes en ET con características demográficas similares. ¿Podría sin embargo suceder que los participantes en ET, voluntarios, estuvieran más motivados, más capacitados para encontrar empleo o tuvieran mejores habilidades interpersonales? Si los voluntarios de ET encontraron trabajo en mayor medida, podría ser por estos factores y no por el programa. Pero si esto es cierto, estos voluntarios deberían tener niveles más elevados de empleo *antes* de entrar en el programa. NIGHTINGALE et al. no encontraron "evidencia alguna" de tal "sesgo en la selección en el efecto global de ET." Demetra Smith NIGHTINGALE et al. *Evaluation of the Massachusetts Employment and training (ET) Program*, Washington, D.C.: The Urban Institute Press, 1991), pp. 151-157.

³¹ Demetra NIGHTINGALE, Lee BAWDEN, y Lynn BURBRIDGE, "Summary of Massachusetts ET Choices Evaluation Design," paper presented at the meeting of the Association for Public Policy Analysis and Management, 29 de octubre de 1987, p. 5.

³² *Ibid.*, p. 6.

³³ Office of Research, Planning and Evaluation, *Request for Proposal*, pp. 2-3.

³⁴ NIGHTINGALE et al., "ET Choices Evaluation Design," p.9.

³⁵ NIGHTINGALE et al. *Evaluation of the Massachusetts Employment and Training (ET) Program*, Washington, D.C.: The Urban Institute Press, 1991), pp. 14, 82,85, 89-90, 93, 97 y 9. El estudio analizó beneficiarios de AFDC que habían participado en "componentes de corta duración" de ET (como búsqueda de empleo) entre julio de 1986 y junio de 1987 y aquellos que habían participado en "componentes de larga duración" (como educación básica y formación ocupacional) entre enero y diciembre de 1986. La duración en AFDC se medía desde que se abría el caso hasta diciembre de 1987. La subvención mensual AFDC se medía entre julio y diciembre de 1987. Empleo e ingresos se medían entre enero y junio de 1988. La vuelta a AFDC se medía hasta diciembre de 1987: "Este resultado... podría haber sido distinto si el período de seguimiento hubiera sido más largo." NIGHTINGALE et al. *Evaluation*, p. 9.

³⁶ *Ibid.*, p. 11

³⁷ *Ibid.*, pp. 10 y 14

³⁸ *Ibid.*, pp. 9, 93 y 90.

³⁹ *Ibid.* 9.

⁴⁰ *Ibid.*, p. 125.