

International Initiative for Impact Evaluation



WORKING PAPER 4
IN SPANISH

Diseño de evaluaciones de impacto: Perspectivas diversas

Robert Chambers, Dean Karlan, Martin Ravallion and Patricia
Rogers

Julio 2009

Acerca de 3ie

The International Initiative for Impact Evaluation (3ie) trabaja con el propósito de ayudar a mejorar las vidas de las personas en países en vías de desarrollo apoyando la producción y utilización de evidencia sobre qué intervenciones funcionan, cuándo, por qué y a qué costo. 3ie es una nueva iniciativa que responde a las demandas de mejores evidencias, la cual mejorará la efectividad del desarrollo promoviendo políticas mejor informadas. 3ie financia evaluaciones de impacto de alta calidad y campañas para mejorar el diseño y la planificación de políticas en los países en desarrollo.

La serie de documentos de trabajo de 3ie abarca tanto temas conceptuales relacionados con la evaluación del impacto, como los hallazgos de estudios específicos o reseñas sintéticas.

El presente documento fue editado por el Dr. Howard White, Director Ejecutivo de 3ie.

© 3ie, 2009

Contactos

International Initiative for Impact Evaluation
c/o Global Development Network
Post Box No. 7510
Vasant Kunj P.O.
New Delhi – 110070, India
Tel: +91-11-2613-9494/6885
www.3ieimpact.org

Contenido

Prólogo

Howard White, Director Ejecutivo, International Initiative for Impact Evaluation (3ie)

Para que los pobres cuenten más: El uso de métodos participativos para la evaluación de impacto

Robert Chambers, Instituto de Estudios del Desarrollo, Universidad de Sussex

Algunas reflexiones sobre el uso de pruebas aleatorias en la evaluación del desarrollo: presentación en la conferencia sobre evaluación en El Cairo

Dean Karlan, Universidad de Yale e Innovaciones de Acción por la pobreza / Afiliado al Laboratorio de Acción por la Pobreza Jameel

Evaluación de tres intervenciones estilizadas

*Martin Ravallion,
Banco Mundial*

Correspondencia entre el diseño de evaluaciones de impacto, la naturaleza de las intervenciones y el propósito de la evaluación

Patricia Rogers, Profesora de Evaluación del Sector Público en CIRCLE (Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation) en el Royal Melbourne Institute of Technology.

Prólogo

Los debates sobre los enfoques en el diseño de la evaluación de impacto parecen haber experimentado un impase en los últimos años. Uno de los objetivos de la conferencia internacional titulada "Perspectivas sobre la Evaluación de Impacto" que tuvo lugar del 29 de marzo al 2 de abril en El Cairo, organizada por 3ie, NONIE, AfrEA y UNICEF, era el de reunir voces diferentes y trabajar en busca de un consenso. Una sesión clave en esta dirección fue una plenaria en la cual se le preguntó a varios expertos con diferentes perspectivas cómo enfocarían ellos la evaluación de tres intervenciones: una transferencia condicional de efectivo, un proyecto de infraestructura y un programa de anti-corrupción. La motivación que dio origen a esta sesión fue el hecho de que los debates se estancan cuando se quedan a nivel conceptual, y que puede alcanzarse un mayor grado de consenso una vez que se pasa a las especificidades del diseño de una evaluación particular. Me siento muy complacido porque los cuatro presentadores hayan aceptado escribir sus puntos de vista de modo que puedan ser difundidos más ampliamente.

Mis agradecimientos a Hugh Waddington y Rizwana Siddiqui por su colaboración en la preparación de esta colección.

*Howard White,
Director Ejecutivo, 3ie*

Para que los pobres cuenten más: El uso de métodos participativos para la evaluación de impacto

Robert Chambers, Instituto de Estudios del Desarrollo, Universidad de Sussex

Resumen

El punto de partida para una evaluación es preguntarse por qué se está llevando a cabo, quién se beneficiará y qué impacto tendrá la evaluación misma y de qué manera. Los enfoques y los métodos participativos encajan dentro de un paradigma pluralista, evolucionario e iterativo, e incluyen cuestiones tales como el análisis de accionistas, el recuento de historias individuales, el mapeo social participativo, la vinculación causal y la diagramación de tendencias y cambio, el puntaje y "lluvia de ideas" sobre las fortalezas y debilidades de los programas. Bien diseñados y aplicados, los métodos participativos son rigurosos, además de que ofrecen valoraciones cualitativas que pueden también contar lo incontable y generar estadísticas de dimensiones pertinentes que de otro modo pasarían inadvertidas o que serían consideradas como puramente cualitativas. Ellos abren los estudios a las voces de los más afectados por un determinado proyecto de un modo tal que resultaría imposible empleando métodos más convencionales, pudiendo lograrse así que las realidades y las experiencias de los pobres cuenten más.

Comentarios introductorios

A pesar de sentirme halagado por la invitación, la incompetencia comparada me hizo dudar de si aceptar formar parte de este panel. La mayor parte de los presentes en esta conferencia ha diseñado evaluaciones [hicimos un sondeo alzando las manos y vimos que era más de la mitad]. Yo sólo he diseñado una: fue en 1970 y fue un desastre. Estuve renuente también porque vivimos un momento de innovaciones explosivas con metodologías participativas incluyendo aquellas para el monitoreo y la evaluación y los tres programas escogidos se prestaban menos a éstas que aquellos que se preocupaban -por ejemplo- por el desarrollo comunitario, la agricultura o el manejo de recursos naturales. También me siento avergonzado, como se han sentido otros oradores, de ser aún otro del norte. Un africano en una de las sesiones dijo: "Cualquier cosa que provenga del norte es verdad".

Reconozcamos que mucha de la 'verdad' – la creatividad y la innovación – con metodologías participativas ha venido y viene del sur, de Asia, de América Latina y notoriamente de África.

No obstante, me animé, por el énfasis que hizo Sulley Gariba en sus comentarios de apertura, acerca de 'empoderar a las comunidades', y el llamado de Erma Manoncourt a 'abrir la puerta a la participación y el empoderamiento de los pueblos'.

En los tres casos considerados aquí –una transferencia condicionada de dinero en efectivo, un caso sobre desarrollo de infraestructura y una comisión anticorrupción– yo defendería un enfoque que fuera pluralista, evolucionario e iterativo. Podrían emplearse métodos mixtos. El punto de partida sería preguntarse acerca de la economía política de la evaluación: Quién ganaría? Quién podría perder? Y cómo? Y especialmente qué intención se tuvo y cómo se anticipó que los hallazgos pudieran marcar una diferencia. Esto bien podría requerir un taller de "lluvia de ideas" con el personal de quien provee los fondos. Si no estuviesen dispuestos o fueran incapaces de encontrar el tiempo para esto, o de ofrecer alguna respuesta a estas preguntas, espero poder tener el arrojo y los recursos para ceder la tarea. Yo negociaría el acuerdo entra las partes para incluir otros pasos. Uno sería el análisis de los participantes y las negociaciones que involucrarían a los agentes interesados en el proceso.

Limitaré mis sugerencias a los métodos participativos que podrían emplearse en lugar de o como complemento de otros. Para éstos, sería prioritaria una búsqueda de buenos e innovadores facilitadores. Yo negociaría el tiempo (probablemente más que el que los entes financiadores hubieran anticipado) para que los talleres y el trabajo de campo evolucionaran y se piloteara una aproximación participativa, evitando un cierre prematuro y -tomando prestado el título del libro escrito por Irene Guijt- "buscando una sorpresa". Por sobre todo, y a lo largo del proceso, habría una pregunta del tipo '¿y ahora qué?' sobre los caminos para impactar la evaluación de impacto misma y su efectividad en términos de costo-beneficio y de valor agregado.

Caso 1: Transferencia Condicional de Efectivo (TCE)

Podría explorarse la aplicación de los siguientes métodos y enfoques participativos:

- Un censo participativo de mapeo en comunidades representativas. Sería importante que se realizara una cuidadosa selección de las comunidades. El mapeo cubriría todas las personas de cada comunidad evitando así cualquier inquietud relacionada con el muestreo. Se facilitaría la identificación de mujeres en edad escolar, escolarizadas y desescolarizadas antes de la aplicación del programa así como aquellas escolarizadas y desescolarizadas en la actualidad. Esto generaría estadísticas que se contrastarían con registros oficiales que incluirían cambios en el desempeño de las niñas.
- Seguimiento de casos individuales, incluyendo casos extremos y personas enojadas, si las hubiere.
- Grupos de enfoque y entrevistas semi-estructuradas como más convenga con niñas, profesores, padres, administradores...
- Invitar a las niñas y a otros a que diseñen y presenten el drama de sus vidas y sus experiencias antes y después.
- Facilitar diagramas de vínculos causales.
- Recolección de historias.
- "Lluvia de ideas" sobre las fortalezas y debilidades del programa y sobre cómo mejorarlo.

Los facilitadores/investigadores podrían permanecer unos cuantos días y noches en las comunidades y encontrarse para compartir sus apuntes. Se esperarían nuevos temas e inquietudes, los cuales se agregarían a la agenda.

Caso 2: Evaluación ex-Post de un programa para el sector de transportes en un país sudasiático

Se exploraría la aplicación de los siguientes métodos y enfoques participativos:

- Grupos de enfoque y entrevistas semi-estructuradas con accionistas clave y personas afectadas, por ejemplo: pequeños empresarios de los sectores informal y formal, contratistas de transporte, conductores, sindicalistas en el puerto y grandes empleadores.
- Para el impacto de carreteras alimentadoras rurales, una selección de un rango de condiciones y comunidades; identificar informalmente beneficiarios y perdedores y luego realizar entrevistas semi-estructuradas y/o grupos de enfoque de las personas afectadas, facilitando: identificación detallada de beneficiarios y perdedores, líneas de tiempo, diagramas de tendencia y cambio, matriz de puntuación de antes y después, diagrama de cambios de vínculos causales y puntaje de los vínculos.

- Estimaciones numéricas de ganancias (y pérdidas) y donde sea conveniente 'interpelar al diagrama' para una valoración más profunda.

La producción de estimaciones numéricas va en contra de la visión de que los enfoques participativos sólo pueden arrojar información cualitativa. No obstante, se ha visto un crecimiento de 'partinúmeros' durante la última década, es decir de enfoques participativos que generan datos cuantitativos (ver Chambers, 2007).

Caso 3: Evaluación de apoyo de donantes a una comisión de anti-corrupción en un país africano

En un campo tan sensible como el de la corrupción, la evaluación sería una intervención por sí misma, lo cual sería ya algo fuera de lo usual, y sería entendida como una oportunidad. Esto se reflejaría a lo largo del proceso para destacar los beneficios y minimizar los daños. Un paso inicial sería identificar las formas y niveles de corrupción (alto nivel, bajo nivel) y los vínculos causales hipotéticos entre el apoyo de parte de donantes y los cambios que puedan haber ocurrido. Las organizaciones no gubernamentales de defensa y derechos humanos junto con los periodistas, serían los informantes claves.

Con aquellos que participen de los tours de estudio y los miembros de la comisión:

- Sustener diálogos reflexivos en grupos de enfoque (pero ver reuniones individuales más abajo)

Con corrupción de bajo nivel, si hubiere cualquier conexión plausible, buscar variadas fuentes de evidencia y valoración:

- Calificación de los ciudadanos o su equivalente, si los hubiere
- Grupos de enfoque de personas molestas
- Conversaciones casuales en cafés, etc.

Con corrupción de alto nivel una gran parte sería:

- Reuniones informales uno a uno en privado, sin tomar apuntes. Para sustentar esto, en la mejor tradición de la inocente pero maquiavélica "facipulación" (facilitación pero con la intención de manipular), a los evaluadores se les suministraría una mesada.

Comentarios finales

Se ha realizado una pregunta sobre cómo contar lo incontable. Los enfoques participativos poseen una habilidad grandemente desconocida para generar números que también sean conmensurables y puedan ser tratados como cualquier otra estadística. Por medio del juicio, la estimación, y la expresión de valores, las personas pueden cuantificar lo cualitativo. El potencial de estos métodos está aún por ser reconocido, como siempre que hay asuntos éticos de por medio. Bien instrumentados, los métodos participativos pueden ser un "gana-gana" – empoderan a las personas a la vez que ofrecen valoraciones creíbles y confiables para quienes diseñan las políticas.

El poder y la economía política de las metodologías han sido traídos a colación como temas de discusión en esta conferencia. He escuchado una inquietud acerca de que las opciones de la intervención pueden verse influenciados por su labilidad para impactar la evaluación por medio de métodos a los cuales se les atribuyen ciertas formas de rigor. Esto no es algo que yo pueda juzgar. Pero, por ejemplo, si este enfoque condujera a inclinarse a favor de programas con inputs simples, estándar y de rápida acción sobre otros que sean más complejos, de largo plazo, pluralistas, participatorios y empoderantes, los costos de oportunidad para los pobres serían altos.

Finalmente, es sorprendente qué tan pocas veces se ha mencionado la primacía y las capacidades de las personas pobres en esta conferencia. El propósito de la evaluación del impacto es aprender y cambiar aquello que hace que la vida sea mejor para ellos. Para lograrlo necesitamos métodos mixtos y pluralismo. Para la valoración del impacto pueden y deberían emplearse muchas aproximaciones y herramientas.

En cualquier caso, éstas deben siempre reconocer que son aquellos que viven en la pobreza, aquellos que son vulnerables, aquellos que son marginados quienes son los mejores jueces y las principales autoridades sobre sus vidas y su forma de ganarse la vida y cómo se han visto afectados. Ahora sabemos -y hace dos décadas no los sabíamos- que ellos tienen mucha más capacidad analítica de lo que pensábamos. Sabemos que 'ellos pueden hacerlo. Para facilitar su propio análisis empoderante contamos ahora con una riqueza de metodologías participativas. Necesitamos hacer un mayor y mejor uso de ellas. Una y otra vez, involucrar a los afectados hace que repitamos: ¡pregúntenles a ellos!

Referencias

Chambers, Robert (200?) 'Who Counts? The Quiet Revolution of Participation and Numbers' IDS Working Paper 296. Falmer: Institute of Development Studies.

Algunos pensamientos sobre el uso de pruebas aleatorias en la evaluación del desarrollo: presentación en la conferencia sobre evaluación en El Cairo

*Dean Karlan¹,
Universidad de Yale e
Innovaciones de Acción
por la pobreza / Afiliado al Laboratorio de Acción por la Pobreza Jameel.*

Resumen

Se nos pidió comentar acerca de algunos enfoques metodológicos específicos para evaluar tres intervenciones hipotéticas. El presente artículo hace uso de este foro para discutir tres errores de percepción acerca de las pruebas aleatorias. En primer lugar, nadie afirma que las pruebas aleatorias sean apropiadas en todas las situaciones y para todas las preguntas. Todos estamos de acuerdo en que es prioritario hacer la pregunta correcta. En segundo, la decisión acerca de qué medir y cómo hacerlo, es decir por medio de métodos cualitativos o participativos versus sondeos cuantitativos o métodos de gestión de datos, es independiente de la decisión acerca de si llevar a cabo una muestra aleatoria o no. En tercer lugar, las pruebas aleatorias pueden ser empleadas para evaluar procesos complejos y dinámicos, no sólo intervenciones simples y estáticas. Los evaluadores deberían enfocarse en resolver las inquietudes más importantes para la toma de decisiones futuras y hacerlo tan confiablemente como sea posible. La confiabilidad se ve mejorada con pruebas aleatorias cuando son viables y si se presta atención a las teorías subyacentes y a las pruebas de por qué las intervenciones funcionan o fallan de modo que las lecciones puedan ser transmitidas de la mejor manera posible a otros escenarios.

1. Introducción

¿Por qué evaluar? Se destacan tres razones: saber dónde invertir recursos limitados, saber cómo mejorar los programas y motivar a aquellos con dinero a dar o invertir más.

Quisiera comenzar con un experimento de pensamiento del filósofo utilitarista Peter Singer. ¿Salvaría usted a un niño que se está ahogando en un lago sabiendo que arruinar su ropa o faltar a una cita le costaría \$100 dólares? La mayoría de las personas responden "sí" a esta pregunta. Pero ¿Enviaría usted \$100 dólares ahora mismo a una ONG en un país pobre para salvar a un niño? Muchos dicen que no, argumentando que nadie sabe con certeza si sus \$100 dólares pueden salvar a un niño o si se van a perder. Esta es una excusa común para la inacción. La evaluación refuta esta excusa.

Ha habido mucha discusión acerca del uso de pruebas aleatorias de control (PAC) versus otros métodos para evaluar programas. No obstante, en muchos casos esta tajante división entre el enfoque experimental y otros enfoques es fabricada y enmascara la sobreposición entre los métodos experimentales y los cualitativos que pueden caracterizar una buena evaluación. En esta nota comienzo por delinear algunos errores comunes de concepción acerca del método de medición, atribución y viabilidad de las pruebas aleatorias de control (PAC). Luego describiré tres ejemplos de programas de desarrollo comunes (transferencias condicionales de dinero, medidas de infraestructura y anti-corrupción) y las circunstancias en las que las PAC deberían o no ser empleadas como parte de la estrategia de evaluación.

¹ Email: dean.karlan@yale.edu.

2. Errores de percepción de las PAC

Un error de percepción común es que uno debe escoger si realizar una evaluación cualitativa o una prueba aleatoria de control. Subyacente a esto hay un espectro erróneo de rigor en la "atribución", con las PAC en un extremo y los métodos cualitativos en el otro. En realidad, las metodologías cualitativas no son contrarias a las PAC. De un lado, una buena evaluación empleando PAC con frecuencia involucra una exhaustiva valoración de cómo funciona el programa, de su diseño inicial, teoría de cambio, participación de beneficiarios, etc.

Para aclarar la discusión sobre métodos de evaluación es imperativo separar nuestras conversaciones acerca de recolección de datos y medición de resultados—qué medir, cómo hacerlo, y a quién incluir en el proceso—desde cómo establecer causalidad entre los resultados y la intervención. Las PAC establecen causalidad ofreciendo una medición de lo contrafactual: ¿qué habría pasado si el programa o la política no hubiesen existido? Así como es práctica estándar en los juicios médicos, esto se logra asignando personas al azar a grupos de tratamiento y control de modo que, excepto por el programa u oferta aleatorios, los grupos sean similares en sus características observables y las no observables si el tamaño de la muestra es lo suficientemente grande.

Estableciendo causalidad

La asignación al azar del programa contribuye evitando un sesgo en la selección, porque los participantes en el programa con frecuencia son distintos a quienes no participan en ellos. Si tuviéramos que comparar a aquellos que podrían participar en un programa pero eligen no hacerlo, terminaríamos comparando a dos grupos de personas muy diferentes. Es fácil ver cómo estos grupos pueden diferir de maneras importantes pero difíciles de medir. Aquellos que se unen al programa pueden estar más inclinados a mejorar su situación, o estar más empoderados o tener una mejor educación. Quizás tienen más tiempo libre. Los investigadores con frecuencia tratan de controlar estas diferencias, pero inevitablemente hay variables que se omiten u otras tales como la motivación, cuya medición puede ser problemática. Estas diferencias significan que las estimaciones del impacto de la intervención pueden estar sesgadas, pues las diferencias en los resultados del tratamiento y los grupos de control puede ser el resultado de estas características inobservadas más que en que hayan sido causadas por la intervención misma.

Datos y mediciones

Las mediciones cuantitativas de resultados son útiles para las evaluaciones porque permiten a los investigadores establecer una significación estadística del impacto del programa. Pero las PAC no especifican un método para la recolección de datos. Tanto los datos cuantitativos como los cualitativos pueden ser empleados dentro del marco de las PAC, con frecuencia combinados con la evaluación misma. Los métodos provenientes de la economía, la sociología y la psicología u otras disciplinas pueden ser empleados, así como procesos participativos que involucren voces locales (por ejemplo, véase Chattopadhyay y Duflo 2004, quienes encontraron que era más probable que las mujeres en Bengala Occidental participaran en el proceso de definición de políticas si el líder del consejo de su aldea para el establecimiento de políticas era una mujer), entre otros, y aún otros participantes externos, tal como discute Chambers en este foro (Karlan y Zinman 2009).

Un error de percepción común de parte de los defensores de los PAC es que sugerimos que pueden y que deberían emplearse para cada programa. Las PAC son una herramienta importante para la investigación porque la causalidad que establecen ofrece una medición rigurosa del impacto del programa y de este modo contribuyen a establecer si deben replicarse en otro lugar, así como en qué manera pueden mejorar. Sin embargo, las PAC no son siempre viables. En dónde sean adecuadas las PAC depende en parte de la situación y también de la pregunta realizada. Como se discute en el artículo de Ravallion (2009) incluido en este foro, uno nunca debería partir de la metodología para luego decidir qué se va a preguntar. Los evaluadores deben establecer primero las preguntas que necesitan respuesta y

luego determinar cuál es la herramienta óptima para responderlas. Siempre que sean viables, las RCT ofrecen la estimación menos sesgada del impacto del programa, pero que sean simplemente viables no indica en modo alguno que deban ser empleadas sólo por utilizar una u otra. En donde no haya estrategias adecuadas para la identificación, las PAC son sin lugar a dudas los medios más prácticos para configurar una investigación confiable.

2.3. Enfoques creativos en las ECA

A la vez que hacemos énfasis en que las PAC no funcionan en todas partes, muchos casos que parecen poco viables se hacen de hecho viables con un poco de creatividad. Por ejemplo, las intervenciones con frecuencia pueden aprovechar los límites de la implementación y aleatoriedad hacia la comunidad u otro nivel geográfico en lugar de seleccionar individuos al azar y distribuirlos en grupos de control y tratamiento. Existen varias evaluaciones que miden el impacto de las micro-finanzas que emplean este enfoque. En otros casos, las diferencias en la intensidad del mercadeo de un programa de diseño por estímulo en diferentes áreas pueden explotarse en una PAC. El criterio clave para las PAC es el tamaño de la muestra, en unidades que puedan ser suficientemente separadas de tal modo que los efectos secundarios y de equilibrio general puedan ser medidos. Si se planean adecuadamente y si los efectos no son resultado de grandes agregaciones (por ejemplo a nivel de país), entonces los diseños cuidadosos de PAC pueden medir tanto los impactos directos de la intervención como los efectos colaterales positivos y negativos hacia grupos distintos a los beneficiarios directos. De hecho, estos son algunos de los más emocionantes PAC que se pueda conocer en la literatura porque nos ayudan a entender no sólo si una idea funciona sobre un individuo en particular, sino cómo los hará a una escala mayor con efectos directos e indirectos.

2.4. Enfoques estáticos de implementación vs. Enfoques dinámicos de implementación

Otro error de percepción común acerca de las PAC es que la intervención debe ser homogénea y estática. De hecho, las intervenciones 'emergentes, complejas' o 'complejas' tales como las discutidas por Rogers (2009) en este foro, no son más difíciles de manejar para una PAC que para una que no sea PAC. Los argumentos que sugieren que la complejidad y un proceso dinámico causan estragos en una PAC fallan al no reconocer qué nos ofrece exactamente una PAC. Una PAC simplemente ayuda a generar un grupo objetivo de comparación contra el cual comparar los cambios. La intervención misma de hecho puede ser estática y simple, o compleja y cambiante. Si es lo último, entonces la evaluación es descrita por supuesto como tal: como una que evalúa un proceso, como una oportunidad asociada a unos recursos, una intervención dinámica y fluida que fue llevada de cierta manera, etc. La clave aquí es que es el proceso, no las actividades individuales que constituyen la implementación del programa, el que es evaluado de esta manera. Si el proyecto consistiera en trabajar, entonces lo que debería replicarse es el proceso de asignación de los recursos, facilitación de su uso, etc. Esto es muy afín a muchas intervenciones en desarrollo comunitario en las cuales los recursos tales como el entrenamiento y la asistencia técnica personalizada son ofrecidas a las comunidades y se llevan a cabo ejercicios de facilitación para ayudar a las comunidades a crecer y prosperar.

Estamos realizando una evaluación así empleando un enfoque con PAC, con herramientas cualitativas y cuantitativas, del Proyecto Contra el Hambre en Ghana, así como de un programa de desarrollo liderado por la comunidad en Sierra Leona. Es importante, por supuesto, comprender que lo que se está evaluando aquí es un proceso colaborativo más que una intervención claramente definida. No es posible saber de entrada qué *inputs* seleccionarán los actores particulares, ni tampoco esperar que el mismo proceso en algún otro lugar produzca las mismas opciones. Así, las lecciones de tal evaluación son acerca de los cambios que uno puede esperar de un proceso tal—no de las escogencias específicas e inversiones que los actores elijan realizar, sino del proceso de facilitar y/o financiar las aldeas a medida que desarrollan el programa ellas mismas. Habiendo dicho eso, si a los funcionarios del programa o a los gerentes les interesara medir los impactos individuales de las actividades que constituyen la intervención, podría diseñarse que una RCT entregara resultados discretos de intervenciones complejas. Esto requeriría variar los componentes de la intervención aleatoria, convirtiéndolos en múltiples grupos de tratamiento. La comparación más probable sería el impacto de un conjunto base de servicios, con o sin la interacción de uno o más componentes adicionales.

Las PAC tienen en este punto una ventaja importante sobre otros métodos, porque pueden evitar los sesgos de selección inherentes a muchos programas sociales, así como a atender los impactos directos de las diferentes actividades en un diseño de tratamiento múltiple. Por ejemplo, si uno condujera una evaluación de la formación en negocios y encontrara un aumento en las ganancias, especialmente entre aquellos a quienes se descubrió que realizaban un mejor manejo contable, ¿sugiere esto que debería promoverse la formación en registros contables? Quizás llevar la contabilidad sea un componente clave del entrenamiento o potencialmente que los mejores empresarios desarrollan naturalmente registros contables. Las PAC pueden desenredar estos asuntos asignando a los participantes para que reciban entrenamiento con o sin un módulo especial de registros contables.

Otro error de comprensión frecuente acerca de las PAC es que miden impactos de una intervención sólo sobre el promedio de la población, haciendo caso omiso de los impactos diferenciales sobre diferentes segmentos de la población. De hecho, dada una muestra suficientemente grande y un plan de muestreo que incluya una variedad de personas que puedan ser elegibles para el programa más amplio, una PAC puede ayudar a identificar grupos para los cuales el programa tiene el mayor impacto y grupos para los cuales el impacto es insignificante o incluso negativo. Por ejemplo, un resultado sorprendente de una medición realizada con PAC sobre el impacto de un programa de formación en el negocio de las microempresas en Perú era que los negocios que no expresaban ningún interés en un entrenamiento adicional de hecho se beneficiaban más del programa (Karlan y Valdivia 2008).

3. Tres ejemplos

El artículo de Ravallion en este foro ofrece un panorama excelente de los tipos de preguntas que uno debe hacerse al comienzo del proceso de evaluación con el fin de definir las metas y el alcance de la evaluación y así definir las preguntas clave de la investigación. Tal como lo discute, dependiendo de la unidad de asignación, la aleatoriedad será viable o no. Estos tres ejemplos ofrecen un excelente espectro justamente de este punto. Discutiré aquí dos planes amplios para cada uno sobre cómo podría evaluarse y luego ideas específicas al interior de cada uno sobre cómo las preguntas acerca de preguntas específicas sobre la implementación pueden resolverse por medio de pruebas aleatorias, aún si la intervención principal emplea otros métodos para evaluar su impacto global. Estas ideas no están ahí en lugar de la evaluación general no-experimental, pero pueden brindar métodos útiles para generar datos precisos y objetivos para ayudar a resolver preguntas futuras sobre la implementación.

Regresando al primer ejemplo, el de transferencias condicionales de dinero en efectivo, el método recomendado para la evaluación del impacto es el de pruebas aleatorias de control, el cual implica una recolección de datos cuantitativa y cualitativa. Estos se han realizados en diversos países. En donde los gobiernos han contado con recursos limitados para implementar a gran escala los programas de TCE, la aleatoriedad ha sido una manera especialmente justa y transparente para distribuir los beneficios de una manera organizada. La investigación reciente ha demostrado que, diseñado adecuadamente, las TCE pueden ser un medio efectivo para alcanzar las metas importantes de las políticas públicas. Sin embargo, la pregunta de cómo resulta mejor implementar estos programas está lejos de haber sido resuelta. Por ejemplo, las preguntas de implementación incluyen qué tan frecuentemente realizar los pagos, si debe considerarse los servicios de ahorro, y si hacer coincidir el pago con gastos en educación².

En el caso del segundo ejemplo, infraestructura, hay diversas opciones para los diseños que sean técnicamente viables pero que requieren un grado variable de compromiso de parte de funcionarios del gobierno que administren los programas. A continuación discutiré la rehabilitación de puertos, carreteras troncales y carreteras subsidiarias rurales. Desafortunadamente, a los evaluadores se les pide con demasiada frecuencia evaluar sólo después de que ya es muy tarde. Sin importar el método empleado, es preferible ante todo organizar la evaluación con antelación, tener objetivos claros y ser incluyente acerca de qué

² Para una buena discusión sobre cómo diseñar ambientes de opciones que ayuden a las personas a elegir de manera ética, véase *Nudge* por Richard Thaler y Cass Sunstein.

y cómo medir mejor los resultados.

La evaluación de la rehabilitación de puertos y carreteras troncales puede involucrar un fuerte enfoque en metodologías de evaluación de procesos. El primer paso será establecer un marco lógico, con objetivos tales como el número de días de espera y la cantidad de días de transporte; el costo del envío y transporte por tierra; el valor de los bienes que van a ser transportados; la cantidad de bienes enviados; y el número de barcos, camiones y carros que entran y salen. Existe un potencial para emplear herramientas econométricas dependiendo de efectos diferenciales sobre las industrias y aranceles, por ejemplo. Esto es simplemente un monitoreo del programa y es importante tanto para la gerencia de la implementación y responsabilidad como por los resultados.

En estos casos las PAC pueden ser empleadas para ayudar a responder a aspectos críticos de la teoría de cambio del programa, pero es poco probable que involucren a la intervención entera. Por ejemplo, una pregunta clave para la rehabilitación de puertos en un país en desarrollo podría ser: ¿Pueden los menores costos en transporte conducir a un mayor crecimiento de la industria en el campo? En este caso, uno podría considerar una PAC que al azar subsidiara los costos de transporte en algunas áreas con el fin de examinar como resultado el cambio en la actividad económica.

Responder a las preguntas claves sobre las políticas sobre el impacto de programas para carreteras secundarias por medio de un enfoque con PAC puede ser técnicamente viable, pero también es probable que requiera un gran compromiso de parte de quienes diseñan las políticas. Este ejemplo trae grandes beneficios en términos de lecciones sobre las políticas para otros países, pero también reconocemos que políticamente podría ser difícil de llevar a cabo. Si hubiera suficientes carreteras, y la geografía y los costos de construcción lo permitieran, habría un potencial para una fase aleatoria en la construcción de carreteras. Imagine un plan a diez años para mejorar o construir carreteras secundarias rurales.

Hacer del orden aleatorio es tanto (a) justo como (b) fácilmente evaluable. Esto podría implementarse incorporando una priorización de carreteras dentro del plan a diez años, si algunas vías son más importantes por razones económicas y geográficas (o políticas). Los planificadores de políticas con una mirada empresarial podrían reconocer que una ventaja de emplear una PAC en este contexto es que evita cualquier favoritismo político al decidir sobre la ordenación. (En otras palabras, las carreteras serían seleccionadas para la primera oleada o la segunda oleada por medio de un proceso deliberativo transparente, luego el orden de construcción de las carreteras dentro de cada oleada sería dado al azar, por lo tanto de manera justa). En este caso, la PAC es una faceta del diseño de evaluación. Puede también involucrar el uso de métodos econométricos, incluyendo enfoques como diferencias en diferencias, antes versus después, o una comparación transversal de lo construido versus lo no construido (por ejemplo, pueblos a 5 millas de la carretera reparada o construida versus aquellos a 5 millas de una carretera sin reparar o construir).

El ejemplo final, las medidas anti-corrupción, no representan el mejor ejemplo de una evaluación de estilo atributivo, aunque ciertamente puede haber algunos resultados medibles del proceso tales como arrestos realizados o políticos retirados de sus cargos. Pero puede aprenderse mucho de cómo y por qué los funcionarios públicos recurren a tácticas ilegales. Es más, esta es un área en donde la transparencia en el enfoque a la evaluación realmente importa! Por ejemplo, las auditorías municipales en Brasil fueron televisadas. El trabajo de Olken (2007) en Indonesia es otro buen ejemplo que nos permite aprender sobre la efectividad relativa de diferentes métodos anti-corrupción a través de una mezcla de recolección de datos de tipo cualitativo (percepción de corrupción y métodos participativos a partir de las asambleas en las aldeas) y cuantitativo (calidad de las carreteras actuales).

4. Conclusión

Una ventaja final del enfoque PAC es la independencia que permite en que uno puede establecer pruebas estadísticas anteriores al evento y luego dejar que los datos hablen por sí mismos acerca de si algo funcionó o no. Finalmente, el objetivo de la evaluación debería ser

ayudar a decidir qué hacer en el futuro. Esto es válido tanto para los donantes que necesitan saber en dónde poner su dinero como para los escépticos que desean ver que los programas funcionen y para quienes los implementan, quienes necesitan saber cómo diseñar mejor sus programas. Una parte del trabajo más emocionante emplea métodos mixtos incorporando métodos cualitativos dentro de pruebas aleatorias y empleando métodos aleatorios para evaluar procesos dinámicos y complejos tales como los programas de desarrollo comunitario. En este trabajo me he enfocado en un par de asuntos claves en el debate en torno a los métodos de evaluación de impacto: la decisión de qué medir versus qué comparar. Mirando estas diferentes cuestiones podemos ver que las PAC se enfocan en lo último y son tan flexibles como para incluir muchos métodos participativos, cualitativos y cuantitativos para lo primero. He intentado también disipar algunos errores comunes de percepción acerca de los extremos del debate. Incluso los proponentes de las PAC no abogan por que se utilicen en todo lugar y para todo programa. Si tuviera que adivinar, yo diría que menos del 1 por ciento de los presupuestos para evaluación se emplean en PAC. Creo que debería usárseles más, pero no un 100 por ciento –ni siquiera un 99 por ciento.

Referencias

Chattopadhyay, Raghavendra, and Esther Duflo. 2004. Women's leadership and policy decisions: evidence from a nationwide randomized experiment in India. *Econometrica*, 72(5), 1409–1443

Karlan, Dean, and Martin Valdivia. 2008. Teaching Entrepreneurship: Impact of Business Training on Microfinance Institutions and Clients. Yale University Economic Growth Center working paper.

Karlan, Dean, and Jonathan Zinman. 2009. Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts. *Review of Financial Studies*.

Olken, Benjamin. 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy* 115: 200-249.

Ravallion, 2009. Evaluating Three Stylized Interventions, *Journal of Development Effectiveness*, forthcoming.

Rogers, Patricia J., 2009, Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation, *Journal of Development Effectiveness*, forthcoming.

Thaler, Richard, and Cass Sunstein, 2008, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.

Evaluación de tres intervenciones estilizadas

*Martin Ravallion,³
Grupo de Investigación para el Desarrollo del
Banco Mundial, Washington DC, 20433, USA*

Resumen

En una sesión de esta conferencia se nos pidió a mí y a otros panelistas discutir los diseños de evaluación para tres intervenciones estilizadas: transferencias de dinero en efectivo condicionales, un programa en el sector de transportes y una comisión de anti-corrupción. Este trabajo recoge mis respuestas y elabora algunos puntos sobre ellas, incluyendo las referencias a la literatura. Comienzo con algunas sugerencias generales acerca de los temas que hay que pensar al inicio de cualquier evaluación. Luego trato de ilustrar estos puntos con relación a las tres intervenciones estilizadas.

1. Introducción

A los participantes en esta sesión se les pidió discutir cómo evaluar las siguientes tres intervenciones:

1. Una transferencia de dinero en efectivo condicional en un país Centroamericano, en la cual los hogares reciben un pago mensual si las mujeres en edad escolar asisten a la escuela y cumplen con requisitos específicos de asistencia y desempeño.
2. Un programa en el sector de transporte en un país sudasiático que incluya la rehabilitación de puertos, rehabilitación de carreteras troncales y nuevas inversiones en carreteras subsidiarias rurales.
3. Una comisión anti-corrupción (CAC) en un país africano. El programa incluye ayudar a desarrollar las directrices, infraestructura, actualización y visitas de estudio. Varios programas similares están siendo implementados en seis países.

Estas son tres intervenciones muy diferentes en muchos sentidos, pero la diferencia más importante en el contexto de la evaluación es el grado al que pueden ser entendidas como programas asignados, queriendo decir con esto que si algunas unidades de observación (hogares, compañías, aldeas, áreas) reciben el programa pero otras no. Una diferencia relacionada es el alcance de efectos colaterales, donde unidades no asignadas se ven afectadas (de manera positiva o negativa) por el programa. El hecho de que un programa sea asignado no quiere decir por supuesto que las unidades no asignadas no se vean afectadas. Los efectos colaterales pueden ser una fuente seria de sesgos en los métodos de evaluación clásicos.⁴

En los dos extremos, un programa de transferencia condicional de efectivo (TCE) es un programa asignado dirigido a hogares específicos con efectos colaterales probablemente modestos, pero una CAC es generalmente amplia en términos económicos, aunque pueda tener algunos aspectos asignados tales como cuando se "apunta" hacia órganos específicos del gobierno o a compañías específicas. Entre estos extremos, el programa en el área de transporte es un híbrido; la mayoría de los componentes son asignados, aunque algunos efectos colaterales pueden ser enormes. Se necesitan diferentes herramientas de evaluación

³ Estas son opiniones del autor y no son atribuibles al Banco Mundial ni a ninguna otra entidad afiliada. Dirección: mravallion@worldbank.org. El autor agradece a los participantes de la conferencia sus comentarios y a Phil Keefer, Norbert Schady, Dominique van de Walle y Howard White por las provechosas conversaciones y comentarios.

⁴ Por ejemplo, véase Chen et al. (2009) en el contexto de respuestas de gasto público local a los programas de desarrollo para áreas pobres en China y Oduor et al. (2009) sobre efectos marginales en el caso del mercadeo social del tratamiento para la malaria en Kenya.

para programas asignados versus programas no-asignados, aunque los principios esenciales de la evaluación —incluyendo la necesidad de evaluar el impacto contra un contrafactual explícito—son los mismos. Un programa asignado facilita las comparaciones basadas en observación, mientras que algunas muestras seleccionadas de unidades no asignadas son empleadas para tratar de inferir lo contrafactual bajo ciertos presupuestos identificadores⁵.

Para cada caso comenzaré con algunas preguntas clave acerca del contexto de la intervención, antes de entrar a discutir los temas de la evaluación para estas intervenciones.

Naturalmente eso dificulta ser muy específico sin saber más acerca de la situación, pero lo intentaré. En primer lugar, presento algunos comentarios generales pertinentes a los tres.

2. Preguntas genéricas

Las preguntas clave que quisiera hacer al inicio de cualquier evaluación son las siguientes:

- Por qué esta intervención? Comprender las razones para la intervención específica es importante para diseñar una evaluación útil, pero también es importante más ampliamente para un buen diseño de políticas (lo cual, después de todo, es el objetivo último de la evaluación). Indagar acerca de la lógica de la intervención podría incluso conducir a una intervención diferente.
- Qué conocemos actualmente acerca de este tipo de intervención y cuáles son las brechas de conocimiento más importantes? Casi siempre existe alguna experiencia pasada pertinente. Al inicio, una buena revisión de evidencias pasadas puede ser reveladora y puede influir sobre el diseño de programas y su implementación tanto como sobre los temas escogidos a donde se quiere que se dirija la evaluación.
- Cuál es el contrafactual pertinente? El contrafactual clásico es la ausencia de un programa, pero este no debe ser necesariamente el contrafactual de mayor interés para quienes hacen las políticas, quienes con frecuencia emplearán los mismos recursos en otro programa. Un programa específico puede parecer que funciona bien enfrentado a la opción de no hacer nada, pero es muy pobre frente a otra alternativa viable. Formalmente, el problema de evaluación no es esencialmente diferente si el contrafactual es una alternativa al programa; en principio podemos repetir el análisis relacionado al contrafactual de no hacer nada para cada alternativa posible y compararlas. Esto no se hace con frecuencia al evaluar los proyectos de desarrollo, pero es más común en relación con temas de cuidado de la salud y pruebas médicas (en donde el grupo de control recibe la intervención existente y el grupo de tratamiento reciba la nueva).
- Cuáles son los resultados deseados? ¿En qué periodo? Cuáles son los resultados potenciales no deseados? Claramente, es muy importante conocer los objetivos y cómo pueden ser traducidos a resultados específicos y medibles (de manera cuantitativa o cualitativa). No resulta menos importante saber sobre qué periodo de tiempo se esperan los resultados (positivos y negativos)⁶.
- Cuáles son los parámetros pertinentes que deben ser evaluados? Las evaluaciones clásicas se enfocan sólo en dos parámetros, a saber el impacto promedio de la intervención sobre las unidades a las cuales se les da la oportunidad de tomarla (el parámetro de intención-tratamiento) y el impacto promedio sobre aquellos que la reciben. Sin embargo, los planificadores de políticas por lo general no se preocupan solamente por estos dos parámetros. Otras cuestiones de interés incluyen: ¿funciona la intervención de la manera en que se pretendía? ¿Qué tipos de personas ganan y

⁵ Para una discusión más detallada de las presunciones y métodos empleados en los programas asignados y referencias a la (extensa) literatura, véase Ravallion (2008).

⁶ Véase King y Behrman (2009) para una útil discusión de este punto.

quiénes pierden? ¿Qué proporción de los participantes se benefician? ¿Qué pasa cuando el programa se implementa a mayor escala? ¿Cómo podría diseñarse de otra manera para mejorar el impacto?

- Cuáles son los mecanismos esperados de transmisión? La evaluación de impacto clásica puede ser como una “caja negra” que nos dice muy poco acerca de cómo un programa impacta o no. Para diseñar una evaluación que pueda arrojar luz dentro de esta caja negra uno debe entender la lógica teórica de un programa—las formas precisas en que se espera que la intervención mejore las vidas de las personas. (Discutiré la “evaluación basada en teoría” más adelante.) Los planificadores de políticas con frecuencia son poco precisos acerca de esos mecanismos, por lo cual los estudios de parte de los evaluadores pueden contribuir a exponer la lógica de la intervención y a veces incluso revelar que dicha lógica sea dudosa.
- Cuáles son las opciones metodológicas viables para esta configuración? Habrá limitaciones relevantes a la evaluación (técnicas, económicas, políticas y éticas) e idealmente estas deben ser identificadas desde el inicio. Otros métodos de evaluación deseables podrían ser inviables en la situación especificada. Dentro de las opciones viables, la elección debería depender de las respuestas a todas las anteriores preguntas más que en las preferencias metodológicas del evaluador. Esto puede parecer obvio, pero de hecho no hace parte de la práctica común. Con mucha frecuencia el evaluador trae su propio conjunto favorito de herramientas para realizar el trabajo y elige preguntas que pueden ser respondidas con dichas herramientas específicas (más que al revés). A veces el conjunto de herramientas admisible es notablemente disperso y el evaluador comienza con un único método preferido (tal como la aleatoriedad, algún método econométrico no experimental o una herramienta cualitativa favorita) y busca preguntas que puedan ser respondidas con este método. Por regla, esto no conduce a las evaluaciones más útiles.

Las respuestas a estas preguntas son con frecuencia la clave para el diseño de una evaluación, pero serán específicas para cada programa y su configuración. No hay una sola manera correcta y que se adecúe a cualquier intervención dada independientemente de la situación. A la luz de esto, permítaseme ofrecer algunas reflexiones sobre los temas de evaluación presentados por estas tres intervenciones bastante diferentes entre sí.

3. Un programa de transferencia condicional de efectivo (TCE)

Este es el más fácil de los tres por dos razones: primero, una TCE es un programa asignado y es probablemente razonable asumir que los efectos colaterales sean mínimos. Segundo (y en parte debido a la primera razón), ha habido una gran cantidad de investigación evaluativa sobre las TCE, así que sabemos bastante acerca de los temas relacionados con estos programas y cómo evaluarlos de la mejor manera. Sin embargo, aun hay brechas en el conocimiento.

La idea esencial de un TCE es que la familia receptora debe demostrar asistencia regular a la escuela y (en algunos casos) a los servicios de salud; el pago sólo se realiza si estas condiciones (llamada en ocasiones ‘co-responsabilidades’) son comprobadas⁷. Algunos ejemplos tempranos e influyentes fueron el programa Food-for-Education Program en Bangladesh, el programa PROGRESA (Programa de Educación, Salud y Alimentación) en México (ahora conocido como Oportunidades) y Bolsa Escola en Brasil. Existe evidencia de evaluaciones de impacto de que tales programas de TCE traen beneficios no negligibles a hogares pobres en términos tanto de ingresos actuales como futuros a través de mayores inversiones en escolaridad infantil y salud; para una reciente revisión de la evidencia de evaluaciones pasadas véase Fiszbein y Schady (2009).

⁷ El término “transferencia condicional de dinero en efectivo” es equivocado, pues la mayoría de los programas de transferencia y protección social imponen condiciones a los beneficiarios, pero el término se ha aplicado únicamente a este tipo específico de programas.

El razonamiento para una TCE no es tan obvio como uno podría pensar. Una pregunta perdurable ha sido por qué se imponen las condiciones en lugar de realizar transferencias incondicionales. Las fallas en el mercado de créditos por las cuales los hogares pobres no pueden realizar préstamos para financiar la escolaridad de sus hijos se citan con frecuencia como la razón de ser de una TCE, pero puede haber mejores políticas para atender ese problema, incluyendo la de transferencias incondicionales (Das et al. 2005). Los asuntos sobre desigualdad intrafamiliar y economía política con frecuencia dominan otros razonamientos incluyendo las fallas en el mercado crediticio.

Una de las brechas de conocimiento clave acerca de estos programas es qué tan bien funcionan en países con ingresos muy bajos (la mayoría de los programas y evaluaciones vienen de países con ingresos medios, notablemente América Latina). Se está cerrando esta brecha en el conocimiento con base en las TCEs que actualmente se evalúan (notoriamente en África) y de las que se esperan muchos nuevos resultados en los próximos años.

Necesitamos saber más acerca de la composición del paquete de instrumentos de una TCE para tratar de determinar cómo fijar la composición precisa de las transferencias (qué tanto y si es en efectivo o en especie) e incentivos para el cambio comportamental (cuáles comportamientos deben fomentarse). Existe alguna evidencia de que un cambio en la asignación del subsidio por matrícula que sea neutral al presupuesto en el caso de PROGRESA, pasando de escuela primaria a secundaria habría arrojado mayores logros escolares al incrementar la proporción de niños que habrían continuado a la secundaria (de Janvry y Sadoulet, 2006). Este es un ejemplo de mi reflexión anterior de que los evaluadores deben comprender mejor precisamente cómo impactan los programas, de modo que podamos aconsejar mejor sobre cómo mejorar su diseño.

Mientras que ha habido algunas buenas investigaciones evaluativas sobre respuestas comportamentales, ésta se ha enfocado más en los comportamientos intencionados por los diseñadores del programa (a saber, cumplimiento con las corresponsabilidades del programa). Necesitamos saber más acerca de otras respuestas, incluyendo el suministro de empleo y la decisión de ahorrar de parte de los padres.

Las investigaciones pasadas también han dejado baches en el conocimiento acerca de los impactos sobre el bienestar por los cambios comportamentales inducidos. Los resultados deberían incluir la pobreza tanto actual como futura. La actual es más fácil de medir (aunque aún no tan fácil de medir). Puede ser útil enfocarse en “resultados intermedios” tales como el trabajo infantil y la asistencia escolar, pero sabemos menos de los resultados finales. Con seguridad más niños de familias pobres van a la escuela primaria (digamos) cuando el precio de la escolaridad es reducido por medio de las condiciones. ¿Pero aprenden algo útil? ¿O es muy tarde para afectar sus habilidades de aprendizaje? ¿Qué hay acerca de los impactos sobre ganancias laborales posteriores a la escolaridad?

La importancia de los factores contextuales relacionados con el suministro aún está por verse en las TCE. Los resultados frustrantes en términos del aprendizaje infantil y estatus nutricional revelados por algunas evaluaciones pasadas pueden vincularse a factores de suministro—pobre calidad de las escuelas y clínicas—que varían de un lugar a otro. Los cambios comportamentales inducidos por una TCE pueden ser insuficientes para alcanzar los resultados de bienestar deseados sin mejoramientos en el suministro (Fiszbein y Schady, 2009).

Las brechas importantes en el conocimiento se relacionan con los impactos a largo plazo de las TCE. Naturalmente, esto tomará tiempo. Pero es necesario hacer las inversiones en información ahora. Las externalidades involucradas y las consecuentes dificultades en la financiación y sostenimiento de estudios de impactos a largo plazo son severas.

Aún otra brecha en nuestro conocimiento sobre las TCE tiene que ver con su flexibilidad para ajustar la elegibilidad a los cambios necesarios. Esto es importante si estos programas

deben servir una función de red de seguridad, tal como ocurre durante la presente crisis financiera global. Uno espera que será más fácil implementar pagos adicionales a los beneficiarios actuales que expandir temporalmente el número de beneficiarios durante una crisis. Pero se requiere más investigación sobre el tema y sobre cómo los programas de TCE pueden hacerse más flexibles en la práctica.

Al pensar acerca de las opciones para la evaluación, es importante mirar con detenimiento los planes de introducción progresiva para el programa de TCE. Aún si el programa no está siendo implementado como base de prueba, no siempre será viable (sobre bases presupuestales y técnicas) para introducir el programa a nivel nacional de una vez. Algunas áreas geográficas pueden recibir el programa antes que otros. Es clave comprender la selección de áreas como objetivos. Buscar opciones para crear un grupo de control a partir de las áreas no participantes, similares por medio de observaciones. O si existe una fórmula para seleccionar qué áreas deben ir primero, uno también podría considerar un diseño discontinuo que identifique el impacto en un vecindario ubicado en el punto de corte de elegibilidad. A medida que se expande el programa, entrarían las áreas de comparación, pero antes podría haber opciones para identificar los impactos. Aquí puede ser importante saber si las áreas comparadas saben que estarán uniéndose al programa y cuándo se espera esto, dado que puede generar efectos de contaminación en donde las unidades de comparación se vean afectadas por la anticipación de unirse al programa.

En muchos sentidos la evaluación original de PROGRESA es un modelo para otras TCE. La característica clave en el diseño es que un grupo 'aleatorio' actuó como el de control en el periodo de introducción del programa. La longevidad de este programa (sobreviviendo a los cambios de gobierno) y su influencia sobre la comunidad en vías de desarrollo claramente se derivan en parte del esfuerzo sustancial y público puesto en su evaluación. Una tercera parte de las comunidades muestreadas consideradas elegibles para el programa fueron escogidas al azar para formar un grupo de control que no recibió el programa durante un periodo inicial durante el cual los otros dos tercios lo recibieron. El acceso público a los datos de la evaluación ha facilitado un número de estudios valiosos. Puede encontrarse un repaso comprensivo del diseño, implementación y los resultados de la evaluación de PROGRESA en Skoufias (2005).

Es posible mejorar el diseño de PROGRESA. Sería deseable un recuento más completo de las variables de resultados, tal como se discutió anteriormente. Existen preocupaciones acerca de la posible contaminación del grupo de control, notoriamente por acción de efectos anticipatorios (especialmente cuando como en PROGRESA, el programa adquiere dimensión nacional sobre el curso del periodo de evaluación y naturalmente tienen un alto perfil público). El diseño de la evaluación también podría haber hecho más para informar sobre asuntos de diseño del programa tales como si el subsidio debería otorgarse al nivel de escolaridad primaria o secundaria. Los factores relacionados con suministro también han emergido como importantes en las investigaciones subsiguientes, a lo cual puede haber contribuido una mejor aplicación de encuestas vinculadas de instalaciones escolares y de salud.

4. Un programa en el sector de transporte

Existe un número de asuntos contextuales importantes que pensar al evaluar los programas en el sector de transportes. Lógicamente, el primer paso es comprender la necesidad de la intervención, la cual se relacionará con (inter alia) la historia económica, la geografía y la economía política. ¿Acaso fueron los puertos específicos y los vínculos de transporte selectivamente negligidos haciendo necesaria su rehabilitación, y si fue así, por qué? También es probable que las preguntas relacionadas con las finanzas públicas y efectos colaterales estén aún por resolverse. ¿Cómo debe financiarse el programa y acaso es posible que existan efectos de fungibilidad o volatilidad (en donde, a pesar de fungibilidad parcial, la ayuda continúe "fijada" al sector para el que fue intencionada)? ¿Estamos evaluando la intervención correcta?

Los impactos pueden esperarse tanto en el corto como en el largo plazo. Habrá unos efectos temporales sobre las ganancias del trabajo (tanto para los trabajadores en los proyectos como para otros empleados que puedan ganar del apretujamiento de los mercados laborales), lo cual puede ser importante en algunos contextos, tales como cuando el proyecto de transporte hace parte de la respuesta a una crisis. Habrá también varios costos en la fase de construcción más allá de los costos de construcción tales como los impactos sociales y ambientales (incluyendo las familias desplazadas).

Pero el grueso de los impactos se espera una vez que la construcción haya concluido y los esfuerzos de evaluación deben planearse de acuerdo con este horizonte de tiempo. Y aquellos impactos pueden ser de amplio rango y difíciles de cuantificar. El ahorro en costos de transporte (tanto los costos personales y el valor del ahorro en tiempo de desplazamiento para personas tanto como para carga) será sin lugar a dudas de máxima importancia. Para los grandes proyectos probablemente también habrá efectos más amplios (de equilibrio general) sobre la actividad económica, incluyendo su geografía (a la cual volveré más adelante).

La evaluación deberá considerar el funcionamiento de y las implicaciones para los mercados. Querremos saber qué tan bien trabajan los mercados existentes para bienes geográficamente intercambiados, lo cual eleva la posibilidad de identificar los impactos por medio de los precios. (Una de las principales formas en que un proyecto de transporte afecta el bienestar es por medio de los cambios en los precios). Deberíamos también explorar qué tan bien funcionan los mercados de la tierra existentes en el contexto específico, lo cual puede ofrecer un margen para identificar impactos a través de la capitalización en los precios de las tierras. (El modelo clásico de Von Thunen predice que las rentas de las tierras declinarán entre más distantes del centro de la ciudad, reflejando los costos de transporte; para un ejemplo de cómo este tipo de modelo puede ser explotado en la evaluación de proyectos de transporte véase Jacoby, 2000.)

Quizás queramos ver también cómo la intervención ha afectado la geografía de la actividad económica, incluyendo el mercado local y el desarrollo institucional. Por ejemplo, ¿puede decirse que el mejoramiento de carreteras rurales atrae a los mercados, instituciones y nuevas actividades económicas hacia áreas pobres y rezagadas, o que fomenta una mayor concentración geográfica de estas actividades? ¿Cómo pueden los grandes proyectos de transporte cambiar la geografía de la actividad económica?

Dada la evidente importancia de los costos de transporte sobre las decisiones de ubicación, puede esperarse que un gran proyecto de transporte cambie el panorama de la actividad económica, beneficiando a algunas áreas y actividades, pero posiblemente conduciendo a la decadencia de otras. La nueva geografía económica predice que reducir los costos de transporte para productos agropecuarios o de otro tipo a través de inversiones en la infraestructura de transporte aumentará la concentración geográfica de otras actividades no agropecuarias en áreas urbanas, explotando las economías de aglomeración (Fujita et al. 2001, Capítulo 7). Una evaluación de un proyecto de carreteras rurales apoyado por el Banco Mundial en Vietnam encontró evidencia de que el proyecto estimuló el desarrollo del mercado local en áreas pobres—particularmente des-concentrando la actividad económica; véase Mu y van de Walle (2008). Los impactos rurales locales sobre la concentración geográfica de una actividad pueden diferir bastante de los impactos regionales, incluyendo las áreas urbanas.

Una pregunta que inevitablemente surgirá es si la evaluación debería enfocarse únicamente en los resultados de una eficiencia económica agregada. Se argumenta a veces que el diseño de políticas de desarrollo no debería preocuparse por aspectos de equidad de la asignación espacial de una actividad y que no debe interferir con la asignación determinada por el mercado al emplear otros instrumentos ('espacialmente-ciegos') para atender las preocupaciones acerca de la equidad, incluyendo las preocupaciones que tienen una dimensión geográfica tales como áreas pobres y rezagadas (véase en World Bank (2009) un argumento que sigue estas líneas). Aplicando esta visión a los programas de transporte deberíamos preocuparnos solamente de su eficiencia, tal como es medida por su impacto sobre el ingreso promedio.

Sin embargo, dudo que muchos planificadores de políticas aceptarían esta visión y por buenas razones. Ellos saben que no tienen el conjunto completo de instrumentos de políticas que se necesitan para tratar con las consideraciones de equidad de una manera que no tenga en cuenta a lo espacial⁸. De hecho, en países pobres la ubicación es una de las dimensiones más ampliamente utilizadas para identificar a los pobres con fines redistributivos, en lugar de una mejor información sobre los niveles de beneficio individuales. Se espera por un tiempo que en los países en desarrollo las evaluaciones pertinentes a las políticas de proyectos de transporte necesitarán considerar aspectos de equidad y también dimensiones no relacionadas con el ingreso (incluyendo por ejemplo los impactos sobre salud infantil a través de mejor acceso materno al cuidado de la salud debido a mejoras en el transporte).

El diseño de evaluaciones para un programa de transporte necesitará distinguir entre los componentes que pueden ser tratados como programas asignados (tales como carreteras locales rurales) de aquellos que probablemente tengan impactos mucho más dispersos espacialmente (carreteras troncales, por ejemplo) y que por ende requerirán un enfoque más amplio a nivel de sector o de economía.

Las evaluaciones de componentes asignados y no asignados requieren diferentes diseños. Para componentes asignados tales como mejoras en carreteras rurales locales desearemos identificar el área de cobertura relevante que abarque con seguridad los probables impactos (van de Walle, 2009). También necesitaremos comprender cómo se determinó la asignación geográfica, y encontrar áreas similares por observación que no reciban las nuevas carreteras rurales o que la inversión en rehabilitación. Se requerirían casi con certeza observaciones longitudinales (para permitir una estimación con diferencias en diferencias con correspondencia; véase Ravallion, 2008 para una discusión extensa y ejemplos). Los evaluadores también tendrán que evaluar las fuentes probables de sesgos en la selección. Si la selección está basada de manera importante en características observables, debería ser posible recolectar la información correcta para corregir adecuadamente este sesgo. Si la selección está basada en factores no observables, será necesario considerar el alcance de métodos estadísticos tales como la estimación de variables instrumentales. (La aleatoriedad raramente será viable por obvias razones).

Si el proyecto de transporte es grande, entonces también lo será su área de cobertura. Entonces podría ser difícil encontrar buenas áreas para la comparación que no hayan sido afectadas por el programa. Necesitaremos entonces recurrir a otras herramientas⁹. En un extremo (en términos de agregación), se han empelado las regresiones en el crecimiento a través de países para intentar identificar los impactos de la infraestructura, incluyendo el transporte; véase por ejemplo Calderón y Servén (2008). La información geográfica sub-nacional vinculada a los datos de hogares y empresas puede ofrecer una mirada más fina; véase por ejemplo Jalan y Ravallion (2002), quienes encontraron que las carreteras rurales en China impactan positivamente sobre el proceso de micro-crecimiento. Los modelos de equilibrio general espacial computable, fundados sobre la Nueva Geografía Económica (NGE), pueden entrar a jugar un importante papel en el futuro (Fujita et al. 2001).

Estos modelos están conceptualmente bien adecuados al problema de evaluar los impactos de grandes proyectos de transporte, aunque también son complejos modelos que son exigentes en términos de sus requerimientos de información y calibración. La Unión Europea ha desarrollado un modelo espacial de CGE para evaluar las mejoras al transporte; para una aplicación en el contexto de evaluar un gran proyecto de transporte (una vía de tren en Holanda) véase Knaap y Oosterhaven (2000).

⁸ Desde hace mucho tiempo ha sido reconocido en economía que debe haber condiciones bastante fuertes para que haya una estricta separación de los instrumentos de equidad y eficiencia para alcanzar objetivos de bienestar social. Se ha reconocido ampliamente que aquellas condiciones (notablemente la viabilidad de transferencias no distorsionables de pagos únicos) no se dan en la práctica.

⁹ Un repaso útil de las opciones metodológicas puede encontrarse en Oosterhaven y Knaap (2003).

Debe reconocerse que los modelos pasados de NEG han tomado algunos presupuestos bastante poco plausibles sobre cómo ellos modelan los costos de transporte, y estas pretensiones bien pueden jugar un papel en las lecciones obtenidas sobre los impactos de las mejoras en el transporte¹⁰. Puede esperarse para el futuro que haya mayores avances en la construcción de modelos operacionales sobre presupuestos más realistas.

5. Una comisión anti-corrupción

Se trata de un cuerpo independiente con una autoridad inusual para investigar y penalizar la corrupción, reportando típicamente a la cabeza del estado o parlamento¹¹. Deberán considerarse dos preguntas contextuales principales en la evaluación de una CAC. Primero, se requiere una evaluación de si la intervención se dirige a las causas conocidas de la corrupción. ¿Están las raíces de la corrupción en los mecanismos de aplicación o en los incentivos que enfrentan los actores políticos para pasar por encima de los funcionarios públicos? ¿Deberíamos estar hablando en cambio de reformas a las instituciones fiduciarias, especialmente en la gerencia del sector público? En segundo lugar, los evaluadores necesitarían tener claro qué es lo que motiva la intervención. ¿Se trata solamente del deseo de parte de los donantes de una acción que castigue a los funcionarios corruptos, o a un líder político capaz de reprimir a la oposición, o acaso la intervención proviene de un análisis más profundo del problema?

Las evaluaciones pueden ser diseñadas para evaluar el impacto sobre la corrupción del tipo de herramientas empleadas por una CAC. Un ejemplo innovador es el de Olken (2007) quien estudió la corrupción en carreteras en Indonesia empleando un diseño aleatorio. Olken encontró que aumentar el número de auditorías gubernamentales reducía la extensión de 'gastos faltantes' tal como los medía la diferencia entre los costos del proyecto oficialmente registrados y un estimativo independiente de parte de los ingenieros. (Como contraste, Olken (2007) encontró que las aproximaciones participativas al monitoreo tuvieron poco impacto). Mientras que tales estudios evaluativos pueden ofrecer inputs útiles, ellos no identifican el impacto de la CAC como tal. El estudio de Olken nos dice que el monitoreo de parte de auditores del gobierno puede ayudar a combatir la corrupción en Indonesia; sin embargo, por lo menos en una evaluación (posiblemente fechada), los esfuerzos de tipo CAC de la propia Indonesia no parecen haber sido efectivos para facilitar tal monitoreo y castigo más allá de algunos bien publicitados arrestos de individuos notoriamente corruptos asociados con el régimen político anterior (Sherlock, 2002).

El típico CAC no es un programa asignado como el TCE o algunos componentes del programa de transporte discutido anteriormente. Puede haber algún alcance para la incorporación progresiva en cuanto a cubrimiento geográfico y/o ministerial, asignando los poderes de la CAC a un gobierno local o a ministerios y no a otros. Si la incorporación progresiva es viable y los efectos de la anticipación probablemente no son muy severos, comparar los resultados de pruebas objetivas de corrupción (tales como los empleados por Olken, 2007) entre agencias de gubernamentales similares por observación (en diferentes áreas de gobierno local), con la diferencia de que algunas estén bajo escrutinio de la CAC y otras no, podría servir para evaluar el impacto de la CAC. Sin embargo, aún aparte de la viabilidad (el gobierno bien puede preferir dar a las CAC rienda suelta desde el comienzo), habría serias preocupaciones acerca de los efectos anticipatorios y sesgos en la selección. (La señal de que habría sesgo en la selección es ambigua sin conocer más acerca de la situación: un Presidente comprometido a luchar contra la corrupción seguramente escogería los lugares o sectores en donde el problema es considerado mayúsculo; un Presidente que emplea la CAC como una herramienta política escogería algo diferente, incluso evitando los focos de corrupción).

¹⁰ Me refiero aquí a la "función iceberg de transporte de costos" la cual implica que los precios entregados de los bienes aumentan exponencialmente con la distancia a que son enviados, lo cual no es coherente con la evidencia; para una discusión más detallada véase McCann (2005).

¹¹ Un repaso útil de la historia y registros de las CAC puede encontrarse en Rose-Ackerman (1999, Capítulo 9).

Una aproximación más prometedora (quizás complementaria) es la de considerar los posibles canales de impacto de la CAC construidos sobre un modelo teórico más o menos explícito de por qué existe la corrupción. El esfuerzo de la evaluación podría entonces enfocarse en los que parecerían ser los vínculos clave en la cadena causal esperada, de manera similar al tipo de "evaluación basada en teoría" discutida en Weiss (2001), Rogers (2009) y White (2009). Por ejemplo, Klitgaard (1988) ha argumentado que la corrupción es el resultado de tres factores: un monopolio sobre algún recurso, el poder discrecional de los funcionarios para permitir el acceso a ese recurso y la ausencia o falla de los mecanismos para hacer responsables a esos funcionarios. Podría entonces comenzarse por evaluar cómo la CAC atiende cada uno de esos elementos cambiando los incentivos que enfrentan los funcionarios en el ambiente institucional específico. Esto requerirá evaluaciones de qué incentivos enfrentan los funcionarios individualmente y cómo eso ha cambiado con la CAC. Eso nunca será fácil, pues puede haber algunos pocos incentivos que revelen la verdad a un evaluador. Pero al menos construyendo el esfuerzo de la evaluación sobre la comprensión de por qué vemos la corrupción en primer lugar (yendo mucho más al fondo que lo que atañe a la "inmoralidad" de los funcionarios) puede tenerse alguna posibilidad de determinar si la CAC está teniendo algún impacto real sobre el problema.

Las CAC son muy amigas de citar el número de funcionarios judicializados como una medida de éxito. Esto puede ser engañoso, dada la posibilidad de que la CAC involucre una selección parcializada contra opositores políticos. Reconociendo esto, la evaluación podría enfocarse de manera más útil directamente en el apoyo político o en la oposición contra la CAC. Podría recogerse información sobre la filiación política de los blancos de la CAC, junto con encuestas de empleados sobre el contacto con políticos y las razones para la salida de empleados despedidos. Se requerirán entonces las encuestas de seguimiento (tras la encuesta de base), preguntando en los hogares, las compañías y a los funcionarios gubernamentales acerca del nivel de politización de la CAC en comparación con otras instituciones tales como el Banco Central, los ministerios de Hacienda y Finanzas y la oficina del Fiscal General.

1. Conclusión

El arte de una buena evaluación es hacer las preguntas correctas desde el comienzo, motivados por los vacíos de conocimiento existentes, y ajustar los datos y el análisis para resolver esas preguntas en el contexto específico. No es posible anticipar las más importantes preguntas para la evaluación específica en un documento como éste, ni anticipar todas las opciones de métodos de evaluación, pues ambos dependen naturalmente del contexto específico. Sin embargo, se tiene la esperanza de que este trabajo haya ofrecido algunos puntos de partida útiles para esta reflexión.

Referencias

Calderon, Cesar and Luis Servén, 2008, Infrastructure and Economic Development in Sub-Saharan Africa, Policy Research Paper 4712, World Bank, Washington DC.

Chen, Shaohua, Ren Mu and Martin Ravallion, 2009, Are There Lasting Impacts of Aid to Poor Areas? Evidence from Rural China, Journal of Public Economics 93: 512-528.

Das, Jishnu, Quy-Toan Do and Berk Ozler, 2005, A Welfare Analysis of Conditional Cash Transfer Schemes, World Bank Research Observer, 20(1): 57-80.

De Janvry, Alain and Elisabeth Sadoulet, 2006, Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality, World Bank Economic Review 20(1): 1-29.

Fiszbein, Ariel and Norbert Schady, 2009, Conditional Cash Transfers for Attacking Present and Future Poverty, World Bank Policy Research Report, World Bank, 2009.

Fujita, Masahisa, Paul Krugman and Anthony Venables, 2001, The Spatial Economy, Cambridge, Mass.: MIT Press.

- Jacoby, Hanan, 2000, Access to Markets and the Benefits of Rural Roads, *Economic Journal* 110: 713-737
- Jalan, Jyotsna and Martin Ravallion, 2002, Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China, *Journal of Applied Econometrics* 17(4): 329-346.
- King, Elizabeth M. and Jere R. Behrman, 2009, Timing and Duration of Exposure in Evaluation of Social Programs, *World Bank Research Observer* 24(1): 55-82.
- Klitgaard, Robert, 1988, *Controlling Corruption*, University of California Press, Berkeley CA.
- Knaap, Thijs and Jan Oosterhaven, 2000, *The Welfare Effects of New Infrastructure: An Economic Geography Approach to Evaluating New Dutch Railway Links*, mimeo Erasmus University Rotterdam.
- McCann, Philip, 2005, Transport Costs and the New Economic Geography, *Journal of Economic Geography* 5: 305-318.
- Mu, Ren and Dominique van de Walle, 2008, *Rural Roads and Local Market Development in Vietnam* Policy Research Working Paper, World Bank.
- Oduor, Jacob, Anne Kamau, and Evan Mathenge, 2009, Evaluating the impact of micro-franchising the distribution of anti-malarial drugs in Kenya on malaria mortality and morbidity, *Journal of Development Effectiveness*, this volume.
- Olken, Benjamin A., 2007, Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy* 115(2): 200-249.
- Oosterhaven Jan and Thijs Knaap, 2003, *Spatial Economic Impacts of Transport Infrastructure Investments*, in *Transport projects, programmes, and policies* (edited by A. D. Pearman, Peter J. Mackie, John Nellthorp) Interdisciplinary Centre for Comparative Research in the Social Sciences, Ashgate Publishers.
- Ravallion, Martin, 2008, *Evaluating Anti-Poverty Programs*. In Paul Schultz and John Strauss. eds., *Handbook of Development Economics Volume 4*, Amsterdam: North-Holland.
- _____, 2009, Evaluation in the Practice of Development, *World Bank Research Observer* 24(1): 29-54.
- Rogers, Patricia J., 2009, Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation, *Journal of Development Effectiveness*, forthcoming.
- Rose-Ackerman, Susan, 1999, *Corruption and Government: Causes, Consequences and Reform*, Cambridge University Press, Cambridge.
- Sherlock, Stephen, 2002, Combating Corruption In Indonesia? The Ombudsman And The Assets Auditing Commission, *Bulletin of Indonesian Economic Studies*, 38(3): 367-83.
- Skoufias, Emmanuel, 2005, *PROGRESA and Its Impact on the Welfare of Rural Households in Mexico*, Research Report 139, International Food Research Institute, Washington DC.
- van de Walle, Dominique, 2009, Impact Evaluation of Rural Road Projects, *Journal of Development Effectiveness*, 1(1): 15-36.
- Weiss, Carol, 2001, *Theory-Based Evaluation: Theories of Change for Poverty Reduction Programs*. in Osvaldo Feinstein and Robert Piccioto, eds, *Evaluation and Poverty Reduction*, New Brunswick, NJ: Transaction Publications.

White, Howard, 2009, Theory-based Impact Evaluation: Principles and Practice, Journal of Development Effectiveness, this volume.

World Bank, 2009, Reshaping Economic Geography, World Bank, Washington DC.

Correspondencia entre el diseño de evaluaciones de impacto, la naturaleza de las intervenciones y el propósito de la evaluación

Patricia Rogers, Profesora de Evaluación del Sector Público en CIRCLE (Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation) en el Royal Melbourne Institute of Technology, Melbourne, Australia¹².

Resumen

El diseño adecuado de una evaluación de impacto requiere una capacidad de respuesta situacional –haciendo corresponder el diseño a las necesidades, limitaciones y oportunidades del caso particular. El diseño debe reflejar la naturaleza de la intervención y los propósitos de la evaluación de impacto. En particular, la evaluación de impacto debe atender los aspectos simples, complicados y complejos de la intervención. Los aspectos simples pueden ser estrechamente especificados y estandarizados; los aspectos complicados funcionan como parte de un paquete causal; los aspectos complejos son adecuadamente dinámicos y adaptativos. Se recomiendan diferentes diseños para cada caso, incluyendo pruebas aleatorias de control, regresión discontinua, entrevistas no estructuradas con la comunidad, informe de historias de desempeño participativo y evaluación de desarrollo.

El enfoque de respuesta situacional a la evaluación de impacto

Esta sesión de la conferencia tenía la intención de demostrar la aplicación de distintos enfoques al diseño de evaluación de impacto. Mientras que las presentaciones se enfocaron particularmente en la medición y análisis causal, vale la pena recordar que existen otras tareas que deben atenderse en una evaluación de impacto (Rogers, 2008a). Un diseño exhaustivo para la evaluación de impacto determina cómo la evaluación va a desempeñar toda la gama de tareas involucradas en la evaluación del impacto, las cuales son: “la identificación exhaustiva de los impactos importantes; la recolección sistemática y sustentable de datos y el análisis de la evidencia de estos impactos; inferencias sólidas acerca de la contribución de la intervención para alcanzar dichos impactos; y el manejo efectivo de la evaluación, incluyendo el informe transparente de la metodología y donde sea conveniente, una meta-evaluación formal” (NONIE Subgroup 2, 2008). Para cada una de estas tareas existe un rango de opciones y el diseño de la evaluación de impacto necesita escoger el método o la combinación de métodos más apropiada para cada tarea.

Algunos enfoques a la evaluación de impacto se centran en el uso de un diseño o método particular para la recolección de datos y su análisis o forma de administración. Mi enfoque sobre el diseño de la evaluación de impacto puede describirse mejor como ‘capacidad de respuesta situacional’ (Patton, 2008a).

Existe un creciente reconocimiento en el desarrollo y otras áreas de la evaluación al hecho de que diferentes situaciones de evaluación se verán mejor atendidas seleccionando adecuadamente un rango de métodos y técnicas. Por ejemplo, NONIE¹³ en su declaración sobre la evaluación del impacto declaró en 2008 que:

“NONIE aboga por un enfoque ecléctico y abierto para encontrar los mejores métodos para la tarea de evaluación de impacto –basado en el amplio rango de técnicas disponibles

¹² Datos de contacto: Patricia.Rogers@rmit.edu.au 124 Latrobe Street Melbourne VIC 3000 Australia.

¹³ NOTA: NONIE es una red de redes para Evaluación de Impacto compuesta por las siguientes entidades: Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC) Evaluation Network, United Nations Evaluation Group (UNEG), Evaluation Cooperation Group (ECG), y la International Organization for Cooperation in Evaluation (IOCE)-una red formada a partir de las asociaciones para la evaluación regional.

provenientes de las diferentes disciplinas." (NONIE, 2008)

Michael Quinn Patton resumió esto en el discurso a que fue invitado a dar sobre "el estado del arte en la medición de la efectividad del desarrollo" dentro del marco de la conferencia del Grupo Independiente de Evaluación del Banco Mundial que tuvo por nombre "Midiendo la efectividad del desarrollo: Progreso y limitaciones", como un nuevo marco de lo que nosotros entendemos como el "estándar de oro" de la evaluación de impacto: "el estándar metodológico de oro aquí es lo que se considere apropiado, appropriateness, no algún método en particular" (Patton, 2008b).

La capacidad de respuesta situacional involucra la correspondencia entre el diseño y las necesidades, limitaciones y oportunidades de la situación particular. Las dos preguntas clave que deben ser respondidas antes de desarrollar un diseño de evaluación de impacto son, por tanto, "¿Cuál es la naturaleza de la intervención?" y "¿Por qué se está realizando una evaluación de impacto?"

¿CUÁL es la naturaleza de la intervención?

El tipo de intervención y su escala son determinantes importantes de un adecuado diseño de una evaluación de impacto. ¿Se trata de un proyecto pequeño que está siendo piloteado para su posible replicación e implementación a mayor escala? ¿Se trata de un programa permanente que probablemente continúe de alguna manera? ¿Es una intervención uniforme o una colección de iniciativas dispares? ¿Está la intervención estrechamente especificada y estandarizada o varía en diferentes locaciones en respuesta a las condiciones, necesidades y oportunidades locales? Estas preguntas tienen implicaciones sobre el tipo de evaluación de impacto que se requiere, la posible disponibilidad de recursos para la evaluación, y las opciones en términos de diseño de investigación.

¿Cuál es la naturaleza de los impactos buscados? ¿Son ellos producidos directamente por la intervención (como un splash) o indirectamente (como una onda)? ¿Son impactos de corto plazo que serán evidentes durante la vida de un proyecto y una evaluación (tal como en el caso del desempeño escolar de los niños) o impactos a largo plazo que serán evidentes sólo muchos años después (tales como el empleo tras terminar la escolaridad o la recuperación del suelo por salinidad)? ¿Se trata de impactos transformativos, los cuales una vez alcanzados son poco probables de reversarse (tales como aprender a leer o montar en bicicleta), o impactos frágiles que pueden ser fácilmente deshechos (tales como una adecuada nutrición o la asistencia femenina a la escuela)? ¿Son los impactos el resultado probable de una intervención de tipo "bala de plata", la cual consigue resultados independientemente del contexto, o una intervención "de tiro en línea" que alcanza resultados sólo en consonancia con circunstancias favorables, incluyendo quizás otras intervenciones?

Estas diferentes características pueden ser resumidas en términos de una tipología tripartita como simple, complicada o compleja (Stacey 1992; Glouberman, 2001; Glouberman y Zimmerman, 2002, Kurtz y Snowden, 2003). Esto ha mostrado ser útil para planear y analizar evaluaciones (Guijt, 2008, Patton, 2008a; Rogers, 2008b). La tipología resulta particularmente útil cuando se la emplea para clasificar los aspectos de las intervenciones más que la totalidad de la intervención.

En esta tipología, el término "complejo" tiene un significado específico e importante, el cual no siempre tiene en su uso común. Respecto de evaluación, el término 'complejo' es empleado a veces como sinónimo de "complicado", a veces empleado para referirse a algo que es difícil, y a veces empleado como una excusa para la planeación inadecuada. En esta tipología, "complejo" se refiere adecuadamente a aspectos dinámicos y emergentes de las intervenciones, las cuales son adaptativas y tienen capacidad de respuesta ante necesidades y oportunidades emergentes. Los aspectos simples de las intervenciones pueden especificarse estrechamente y son estandarizados - por ejemplo, un producto, técnica o proceso específico. Los aspectos complicados de las intervenciones tienen múltiples componentes, son parte de una intervención más grande y multi-componente, o funcionan de manera diferente como

parte de un paquete causal más grande, por ejemplo en ambientes de implementación particulares, para tipos de participantes particulares o en conjunto con otra intervención.

Estos diferentes aspectos de las intervenciones tienen implicaciones significativas sobre cómo operan las intervenciones, cómo podemos comprenderlas y cómo podemos usar esta comprensión, tal como se indica en la Tabla 1.

Tabla 1: Implicaciones de los aspectos simples, complicados y complejos de las Intervenciones:

Aspectos	Implementación de la intervención	Implicaciones para:	
		Procesos causales	Reportes y uso de los hallazgos de la evaluación del impacto
Simples	Una sola organización	Una sola línea causal requerida para producir los impactos	Un solo mensaje: qué funciona
Complicados (múltiples componentes)	Múltiples organizaciones en una relación contractual con roles claramente definidos	Múltiples líneas causales requeridos para producir los impactos: : Múltiples intervenciones secuenciales O intervenciones múltiples simultáneas O múltiples niveles de intervención O diferentes mecanismos causales que operan en diferentes contextos	Mensaje contingente – qué funciona para quiénes y en qué situaciones
Complejos (dinámicos y emergentes)	Múltiples organizaciones en una relación de socios para el desarrollo	La causalidad es recursiva, con bucles de realimentación Resultados emergentes: el todo es más que la suma de las partes	Mensaje dinámico, mensaje emergente: qué está funcionando

¿PARA QUÉ se está llevando a cabo una evaluación de impacto?

El propósito de una evaluación de impacto también debe ser tenido en cuenta al desarrollar un diseño de evaluación. ¿Quiénes son los usuarios previstos de la evaluación? ¿Qué considerarán ellos como evidencia creíble en términos de los impactos que deben ser incluidos, las medidas que debe usarse y la aproximación al análisis causal? ¿A quién debe involucrarse en decidir los parámetros de la evaluación?

¿De quién son los valores que debe emplearse en la evaluación? ¿Cuáles van a ser considerados como impactos significativos, ya sean positivos o negativos? ¿Cuáles serán consideradas como distribuciones deseables de costos y beneficios? ¿El enfoque será sobre el promedio de los efectos o del efecto sobre los menos favorecidos? ¿Cuáles son los usos

intencionados de la evaluación? ¿Se está haciendo para justificar los gastos retrospectivamente, en cuyo caso las estimaciones creíbles del beneficio neto serán suficientes? ¿O se está realizando para informar sobre posibles implementaciones a mayor escala de un proyecto piloto, en cuyo caso se necesitaría buena información acerca de cómo funciona? ¿Se pretende informar acerca del cambio incremental o sobre una reimplantación significativa de un programa o política?

Finalmente, antes de diseñar una evaluación de impacto, debe atenderse ciertos asuntos logísticos. ¿Para cuándo se necesita tener el informe? ¿Qué evidencia está ya disponible acerca de esta intervención y sobre otras intervenciones similares? ¿Qué recursos adicionales están disponibles para realizar la evaluación del impacto?

Siguiendo El enfoque de capacidad de respuesta situacional, es sólo cuando hemos atendido todas estas preguntas que podemos enfrentar la de “¿cómo debería hacerse?” Claramente, en una evaluación real todas estas preguntas podrían ser resueltas antes o durante el proceso de desarrollar un diseño. En este taller de diseño hemos cerrado las brechas en las descripciones de los casos para producir escenarios más específicos.

1. Transferencias condicionales de dinero en efectivo

Este caso se describió como sigue:

Una transferencia condicional de dinero en efectivo en un país Centroamericano, en la cual los hogares reciben un pago mensual si las mujeres en edad escolar asisten a la escuela y cumplen con determinados requerimientos de asistencia y desempeño.

Los programas de transferencia condicional de dinero (TCE) tienen algunos aspectos importantes que se caracterizan mejor como simples – es decir, son intervenciones discretas y estandarizadas cuya intención es ser implementadas de la misma forma en diferentes localizaciones. Sin embargo, las evaluaciones de TECs han revelado una considerable heterogeneidad de resultados (por ejemplo, Soares et al, 2009). Una TCE puede necesitar un paquete de otras intervenciones para alcanzar los resultados pretendidos. Por ejemplo, una TCE puede ser exitosa en alcanzar el objetivo inicial de asistir a la escuela, pero las mejoras en impactos en variables a más largo plazo como el aprendizaje de parte del estudiante, su graduación y resultados laborales dependerán también de haya escuelas efectivas en el sitio. Sería importante considerar entonces los aspectos complicados de la intervención e incluir mediciones y análisis de otros elementos (tales como las características particulares del ambiente de implementación y las características de los participantes) requeridas para alcanzar los impactos esperados.

En este caso, asumiendo que el propósito de esta evaluación de impacto es decidir si implementar a mayor escala un programa piloto, una prueba de control aleatorio podría ser un diseño adecuado para el análisis causal, particularmente si se combina con una recolección sistemática de datos acerca de otros factores tales como la calidad de la escolaridad que puedan plausiblemente requerirse para alcanzar los impactos esperados, y los análisis tales como la desagregación para identificar su contribución. Desarrollar y probar la teoría de un programa que incluya factores adicionales y efectos diferenciales así como resultados intermedios mejoraría la calidad del análisis. Si las TECs son efectivas para mejorar el desempeño escolar pero sólo en combinación con escuelas efectivas, entonces no serán una solución pasajera por sí mismas. De manera alterna, dado que la elegibilidad para una TECs está determinada por un cierto nivel de ingreso, una aproximación basada en regresión discontinua podría suministrar evidencia contundente de una atribución causal a menor costo que un ECA, aunque también necesitaría ser desagregada por entornos de implementación.

Además de medir los impactos esperados, es importante que la evaluación de impacto incluya otros impactos significativos, siguiendo la definición de impacto del CAD (Comité de Asistencia para el Desarrollo) a saber: “los efectos positivos y negativos, primarios y secundarios de largo plazo producidos por una intervención para el desarrollo, directa o

indirectamente, con intención o sin ella”.

Algunos impactos potenciales no intencionados pueden anticiparse y ser incluidos en la recolección de datos enfocada empleando una “teoría negativa de programa” (Weiss, 1997). Es más probable que los impactos no anticipados, particularmente los impactos negativos, sean observados a través de métodos participativos (Chambers, 2009). Entonces pueden necesitarse estudios de caso, basados en iteraciones de entrevistas, observaciones y revisión de documentos, para revelar los impactos no intencionados (positivos o negativos), seguidos de sondeos a gran escala o revisión de estadísticas administrativas dirigidas a la producción de estimaciones de la frecuencia de dichos impactos.

2. Programa para la infraestructura de transporte

Este caso fue descrito así:

Una evaluación ex-post de un programa para el sector de transporte en un país sud-asiático que incluya rehabilitaciones portuarias, rehabilitación de carreteras troncales y nuevas inversiones en carreteras rurales secundarias.

Para este caso asumimos que el propósito de la evaluación de impacto era comprender el impacto global de la inversión principalmente para reportar a los inversionistas y también al público, y que ninguna evaluación había sido diseñada al comienzo del programa. Esta carencia, combinada con la naturaleza de la intervención (la cual es multifacética, diversa y que afecta una región entera) presenta considerables retos para el análisis causal, de manera tal que probablemente sería imposible implementar enfoques clásicos experimentales y cuasi-experimentales efectivamente (Ravaillon, 2009).

La evaluación de impacto en este caso requiere de una forma de recolectar evidencia acerca de un gran número de diversos componentes y luego atender el asunto de la atribución sin establecer una comparación ni tener un grupo de control. En este caso sería apropiado entonces emplear un nuevo enfoque, informe de historias de desempeño participativas, expuesta en esta conferencia (Dart, 2009) la cual puede realizar estas funciones. Dado que para algunas personas el término “participativo” implica una aproximación menos rigurosa a la recolección y análisis de datos, es importante comprender la muy fuerte base empírica que subyace a este enfoque.

PPSR es un desarrollo y sistematización de múltiples líneas y niveles de evidencia (MLNE), una aproximación al análisis causal diseñado para situaciones de “altas apuestas” en donde la construcción de un grupo de control o de comparación no es posible, pero en donde hay disponible una considerable cantidad de evidencia de alta calidad, y donde se requiere el análisis sistemático causal.

“Una línea de evidencia es un tipo de evidencia tal como un atributo de un ecosistema (por ejemplo la abundancia de peces, riqueza de especies macro-invertebradas, biomasa macro-fítica), que es investigada en relación con un causante de stress o una intervención; y un nivel de evidencia es un valor para la fuerza de la evidencia empleado para determinar el caso al inferir que una actividad humana dada causa un cambio ecológico o geomorfológico determinado.” (Cottingham et al. 2005)

Mientras que este enfoque no aparece en la mayoría de las guías acerca de la evaluación del impacto, su uso se remonta a las estrategias empleadas en los análisis de evidencia realizados en los años 60 acerca de la relación entre fumar y el cáncer de pulmón. En ese entonces era importante poder llegar a conclusiones sobre el impacto de fumar sobre la salud, pese a tener considerables datos científicos frente a una ausencia de evidencias del tipo De las ECA. Más recientemente, las MLNE han sido empleadas en evaluaciones de riesgos humanos y ecológicos, así como en la gestión de recursos naturales (por ejemplo, ver Keough et al, 2002, Boyes 2006).

El MLNE involucra investigar sistemáticamente la fuerza del argumento causal que vincula una intervención o una causa y sus efectos por medio del análisis de una asociación

inobservada en términos de un criterio causal particular y por medio de la identificación y descarte de posibles explicaciones alternativas. Para la asociación entre fumar y el cáncer de pulmón, Hill (1965) consideró su fuerza, consistencia, especificidad, temporalidad, coherencia con otras evidencias aceptadas, plausibilidad y la analogía con intervenciones similares. Para investigar las respuestas ecológicas a la intervención gerencial, Norris et al (2005) consideraron la plausibilidad biológica, la respuesta biológica, dosis y coherencia en la asociación. Dada la naturaleza especializada y con frecuencia transdisciplinaria de la evidencia científica, esta investigación fue asumida por un grupo de expertos creíble que abarcaba un rango de disciplinas pertinentes a quienes se les pedía juzgar la credibilidad de la evidencia y el análisis causal. Los reportes de historias participativas de desempeño extiende el MLNE para incluir también la recolección de datos primarios y una reunión de las partes involucradas en donde la información existente, los datos primarios adicionales y el análisis del panel de expertos son revisados para producir un informe de evaluación de impacto acordado, el cual incluye un sendero de evidencia auditable (por ejemplo, Clear Horizon et al, 2008).

La aplicación de PPSR al caso de infraestructura de transporte comenzaría con desarrollar una cadena de resultados o teoría del programa, identificando los impactos esperados y una cadena de resultados intermedios que vinculen el programa a dichos impactos. En segundo lugar, se ubicaría la evidencia disponible acerca de la implementación del programa, el logro de resultados intermedios e impactos finales, y la contribución del programa a estos. Luego se evaluaría la calidad de esta evidencia en términos de su calidad en relación con el tipo de información, no en términos de una jerarquía de evidencia basada en el tipo de diseño de investigación empleado. Por ejemplo, las entrevistas con usuarios del transporte serían evaluadas en términos de los métodos de muestreo empleados, las preguntas realizadas, los procesos empleados para reducir el sesgo, y la posibilidad de auditar la información. En tercer lugar, se recolectaría evidencia adicional para atender vacíos específicos en la información disponible. Por ejemplo, si la información disponible estuviera en su mayor parte en la forma de estadísticas oficiales tales como la cantidad y el tipo de carga enviada, el número de vehículos que utilizan las carreteras y hacen parte del grueso de la actividad económica, la recolección de información adicional podría ser útil para esbozar el efecto del programa de transporte sobre los individuos, hogares y aldeas. Esta recolección de información adicional tendría también la intención de capturar información de un rango de perspectivas acerca de los efectos importantes no esperados, bien sean positivos o negativos. Por ejemplo, el mejoramiento de carreteras locales puede aumentar la accesibilidad a instalaciones médicas centrales, mejorando los resultados de complicaciones en el nacimiento. Podría hacerse seguimiento de ejemplos de los impactos que surgen de un pequeño número de esbozos por medio del examen de estadísticas oficiales de salud.

Luego, se asignaría a un grupo de expertos en un rango de áreas pertinentes para que revisara la información en términos de la credibilidad de sus reclamaciones tanto en la descripción de los impactos como en su atribución (bien sea independientemente o en combinación con otros factores) al programa. Mientras que el enfoque con PPSR no vincula específicamente la revisión de expertos con la recolección de datos adicionales, sería deseable que los datos adicionales estuviesen incluidos en esta revisión y que el panel recomiende aún otra recolección de datos y su análisis para atender los vacíos en la base de la evidencia. Finalmente, una reunión que incluya a los representantes de todos los grupos principales de cada una de las partes sería convocada para revisar el informe de la evaluación de impacto. El reporte final sería publicado en diferentes versiones – un resumen ejecutivo, con los hallazgos más importantes y esbozos que ofrezcan ejemplos del impacto, y un informe detallado con vínculos a toda la evidencia sobre la cual se basaron las conclusiones. El PPSR, por tanto, ofrecería una contabilidad más exhaustiva de las diferentes partes, incluyendo a la comunidad, de los diferentes impactos del programa de transporte que si emplearan otras aproximaciones.

3. Programa anti-corrupción

Este caso fue descrito de la siguiente manera:

Apoyo de donantes a una comisión anti-corrupción en un país africano. El programa incluye ayudar a desarrollar las directrices, actualización de infraestructura y jornadas de estudio. Se están implementando programas similares en seis países.

No se trata de una intervención estandarizada, ni parece ser una que esté estrechamente prescrita con antelación. En cambio, es probable que los objetivos específicos del programa y los medios para alcanzarlos emerjan a medida que avance el programa y se desarrolle una mejor comprensión de sus prioridades y posibilidades. Por esta razón, yo sugeriría el uso de la evaluación “del desarrollo” (Patton, 1994, 2008), la cual no tiene la intención de suministrar un informe al final de la implementación de una intervención estandarizada y fija, sino de proveer información durante la implementación de una intervención continuamente cambiante con importantes aspectos complejos.

« La evaluación del desarrollo se refiere a relaciones de asociación de largo plazo entre los evaluadores y aquellos involucrados en iniciativas innovadoras y desarrollo. Los procesos de evaluación del desarrollo incluyen formular preguntas evaluativas y recolectar información para suministrar realimentación y apoyo a la toma de decisiones con base en el desarrollo así como para realizar correcciones por el camino emergente. El evaluador hace parte de un equipo cuyos miembros colaboran entre sí para conceptualizar, diseñar y probar nuevos enfoques en un proceso de mejoramiento continuo, adaptación y cambio intencional de largo plazo. La principal función del evaluador en el equipo es elucidar las discusiones del equipo con preguntas, datos y lógicas evaluativas, y facilitar evaluaciones basadas en datos de dónde están las cosas, cómo se están desarrollando, qué direcciones son prometedoras, qué direcciones deben ser abandonadas, qué nuevos experimentos deberían intentarse – en otras palabras, toma de decisiones con base en información en los procesos de innovación emergentes y de desarrollo. » (Westley et al, 2006).

En este caso, la evaluación de impacto implicaría trabajar con los implementadores del programa y los participantes para desarrollar una teoría del programa que represente lo que ellos entienden como el mayor de los problemas en términos de corrupción y cómo se plantearon atenderlo. Cada iniciativa desarrollada como parte del programa sería revisada en el tiempo, junto con las medidas adecuadas y las formas de probar la atribución causal. Por ejemplo, si una iniciativa enfocada en una “línea caliente” para que el público informe de manera anónima sobre los casos de corrupción, la evaluación del impacto debería hacer seguimiento de estos casos (lo cual podría involucrar todos los casos, o una muestra estratificada tomada al azar, dependiendo de los números y los recursos disponibles) para ver qué pasó en términos de dónde fueron referidos, qué investigaciones se hicieron, si el reclamo fue sustentado, y si lo fue cuáles fueron las consecuencias para el perpetrador.

Desarrollar teorías de programa para cada iniciativa haría posible identificar diferentes teorías de cambio que puedan ser evidentes – por ejemplo: ¿se pretendía que las intervenciones funcionaran por medio del mejoramiento de la comprensión por parte de los oficiales de los requerimientos de probidad por medio de la identificación, castigo y destitución de funcionarios corruptos, a través de erradicar la corrupción incrementando el riesgo de detección por medio de auditorías formales o haciendo más posible que el público pueda informar sobre incidentes de corrupción? ¿Acaso las teorías particulares de cambio aplican a los diferentes tipos de prácticas corruptas o corrupción en diferentes tipos de programas?

Si un propósito principal de la evaluación era identificar buenas prácticas y traducirlas a otros escenarios, un enfoque de “desviación positiva” podría ser efectivo. Esto involucra identificar lugares o casos en los cuales se estén alcanzando resultados extraordinarios, verificarlos y analizar qué es lo que produce los resultados. La característica significativa de la “desviación positiva” es que esta investigación la realizan aquellos quienes buscan aprender de las buenas prácticas. No se hace por medio de un equipo de evaluación y luego es difundida entre los usuarios de la información. La desviación positiva ha sido empleada en salud pública, nutrición, ablación genital femenina, educación y desarrollo de la agricultura, con un ejemplo reportado en el área de la corrupción, específicamente demandas de extorsión a funcionarios públicos (Horowitz, 2006).

Uno de los retos al evaluar los programas de anti-corrupción es el efecto de la corrupción misma sobre la habilidad de recoger evidencia precisa y tomar acción adecuada como consecuencia de los hallazgos. En este caso, sería necesario prestar atención específica al manejo de la evaluación con el fin de asegurar su independencia y la seguridad de los investigadores. Esto podría incluir una línea directa de reportes a un nivel de confianza en el gobierno, con el fin de evitar que los hallazgos sean escondidos o modificados, y/o la participación de las organizaciones de defensa de los ciudadanos para recolectar o investigar los reclamos de corrupción.

4. Conclusión

De manera interesante, cada uno de estos casos incluiría algún componente de una evaluación basada en teorías, aunque la teoría de programas sería empleada de maneras muy distintas. En el primer caso se usó para identificar los resultados inmediatos e intermedios que podrían indicar el logro de impactos de mayor término, e identificar los factores contextuales que deberían ser investigados en el análisis. En el segundo caso, es utilizado como un marco conceptual para reunir diversas evidencias acerca de un conjunto de componentes diverso. En el tercer caso, se emplea como un marco conceptual para orientar un diseño evolutivo para recolectar y analizar los datos e informar del cambio constante.

El análisis de los casos y sus implicaciones para la evaluación del impacto se resumen a continuación en la Tabla 2. Mientras que esta tabla se enfoca en los aspectos clave de las intervenciones, es probable que todos tengan en algún grado algo de los otros aspectos (por ejemplo, todos los casos muy probablemente tendrían aspectos simples).

Tabla 2 Análisis de casos en términos de características y propósitos

	Aspectos clave de la intervención	Propósitos de la evaluación
Caso 1 – TCE	Aspectos simples – Intervención estandarizada Aspectos complejos – trabaja en conjunto con otros factores y programas	Saber si funciona – y en qué contextos – para informar sobre la política en desarrollo
Caso 2 – Infraestructura de Transporte	Múltiples elementos complicados – diversos que deben funcionar juntos efectivamente	Evaluar los impactos globales de un programa completado
Caso 3 Anti-corrupción	Intervención compleja – adaptativa y emergente, que responda a las necesidades, problemas y oportunidades	Comprender y mejorar un programa en curso y cambiante

Los diseños desarrollados para estos tres casos han demostrado que la capacidad de respuesta situacional requiere un conocimiento de un amplio rango de métodos y técnicas. Aún si para la realización de la evaluación de impacto se contratase a un evaluador externo con pertinente conocimiento y experiencia, el agente que lo comisione tendría que tener una comprensión suficiente del método que le permitiera desarrollar los términos de referencia apropiados, seleccionar un consultor adecuado y gestionar efectivamente el contrato, incluyendo la evaluación de la calidad de trabajo. Las implicaciones de esto para construir la capacidad de evaluación son que necesitamos tanto profundidad (especialistas en métodos particulares) como amplitud (comprensión de que existe una gama de métodos, y cuándo estos pueden ser más adecuados).

Agradecimientos

Gracias a Michael Patton, Kaye Stevens, Howard White y Bob Williams por sus útiles comentarios sobre una versión anterior de este trabajo.

Referencias

Boyes, B. (2006). Determining and managing environmental flows for the Shoalhaven River, Report 1 - Environmental Flows Knowledge Review. NSW Department of Natural Resources, May 2006. Retrieved 14 May 2009 from www.dwe.nsw.gov.au/water/pdf/monitor_sholahaven_sh003.pdf

Chambers, R. (2009) Participatory methods. Journal of Development Effectiveness, this issue.

Clear Horizon and O'Connor NRM (2008) Performance Story Report: A study of the Mount Lofty Ranges Southern Emu-Wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Natural Resources Management region. Canberra: Commonwealth of Australia. Retrieved 16 May 2009 from <http://www.nrm.gov.au/publications/books/pubs/psrmount-lofty.pdf>

Dart, J. (2008) 'Report on outcomes and get everyone involved: The Participatory Performance Story Reporting Technique'. Paper presented at the 2008 Australasian Evaluation Society conference, Perth. Retrieved 16 May 2009 from <http://www.aes.asn.au/conferences/2008/papers/p100.pdf>

Dart, J. (2009) 'Participatory Performance Story Reporting Technique'. Paper presented at the 2009 Impact Evaluation conference, Cairo.

Downes, B. J. L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J., Keough, P. S. Lake, B. D. Mapstone and G. P. Quinn (2002) Monitoring Ecological Impacts: Concepts and Practice in Flowing Waters Cambridge: Cambridge University Press.

Glouberman, S. (2001) 'Towards a New Perspective on Health Policy', CPRN Study No. H/03, Canadian Policy Research Networks Inc., Ottawa.

Glouberman, S. and B. Zimmerman (2002) Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Commission on the Future of Health Care in Canada, Discussion Paper 8. Retrieved 14 May 2009 from http://www.hc-sc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman_E.pdf

Guijt, I (2008) 'Seeking surprise: rethinking monitoring for collective learning in rural resource management'. PhD thesis. Wageningen, the Netherlands.

Hill, A.B. (1965) 'The Environment and Disease: Association or Causation', Proceedings of the Royal Society of Medicine 1965 May;58:295-300.

Horowitz, B. (2006) Bridge Ogres, Little Fishes and Positive Deviants: One-on-one deterrence of Public Functionary Extortion Demands. Retrieved 14 May 2009 from http://www.positivedeviance.org/projects/law/Bridges_final.doc

Kurtz C, Snowden D. (2003) The new dynamics of strategy: Sense-making in a complex and complicated world. IBM Systems Journal 2003;42(3):462-483.

Land and Water Australia () Improving the Natural Resource Management System for Regions. Canberra: Australian Government Publishing Service. Retrieved 14 May from <http://www.rkrk.net.au/images/3/34/PR061220.pdf>

NONIE Subgroup 2 (Network of Networks on Impact Evaluation) NONIE Impact Evaluation Guidance. Retrieved 14 May 2009 from http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf

NONIE (Network of Networks on Impact Evaluation) Statement on Impact Evaluation. Retrieved 14 May 2009 from http://www.worldbank.org/ieg/nonie/docs/IE_statement.doc

Norris, R.; Liston, P.; Mugodo, J.; Nichols, S. (2005) Multiple Lines and Levels of Evidence for Detecting Ecological Responses to Management Intervention. Paper presented at the American Geophysical Union, Spring Meeting 2005.

Patton, M. Q. (1994) 'Developmental Evaluation', *Evaluation Practice*, Vol 15, No. 3: 311-319.

Patton, M.Q. (2008a) *Utilization Focused Evaluation*, 4th ed. Text. Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (2008b). State of the Art in Measuring Development Assistance. Address to the World Bank Independent Evaluation Group. Conference, 10 April 2008, Washington, DC. Retrieved 14 May 2009 from

<http://www.worldbank.org/ieg/conference/results/patton.pdf>

Ravaillon, M. (2009) Evaluating Three Stylized Interventions. *Journal of Development Effectiveness*, this issue.

Rogers, P.J. (2008a) Four key tasks in impact assessment of complex interventions, Keynote address. Workshop on Rethinking Impact. Understanding the Complexity of Poverty and Change, Consultative Group on International Agricultural Research (CGIAR) Cali-Colombia. Cali, Colombia. Retrieved 14 May 2009 from

<http://www.prgaprogram.org/riw/files/papers/Rogers%20material%20for%20workshop.ppt>

Rogers, P.J. (2008b) 'Using programme theory for complicated and complex programmes' *Evaluation: the international journal of theory, research and practice*. 14 (1): 29-48.

Soares, F.V, R. Perez, G.I. Hirata (2009) 'Achievements and Shortfalls of Conditional Cash Transfers: Impact Evaluation of Paraguay's Tekoporã Programme'. Paper presented at the 2009 Impact Evaluation Conference, Cairo.

Stacey, R. (1992). *Managing the Unknowable*. San Francisco: Jossey-Bass.

Weiss, C. (1997). *Evaluation: Methods for Studying Programs and Policies*. (2nd ed.) Upper Saddle River, NJ: Prentice Hall.

Westley, F., B. Zimmerman, M.Q. Patton (2006) *Getting to Maybe: How the World Is Changed*. Random House Canada. Extract retrieved 14 May 2009 from http://innovationlabs.com/r3p_public/rtr3/pre/pre-read/Patton.Developmental%20Evaluation.pdf