

DIRECTIVAS DE DESARROLLO

Evaluación del impacto de los proyectos de desarrollo en la pobreza

Manual para profesionales

Judy L. Baker

*Banco Mundial
Washington, D.C.*



© 2000 Banco Internacional de Reconstrucción
y Fomento/BANCO MUNDIAL
1818 H Street, N.W.
Washington, D.C. 20433

Todos los derechos reservados
Fabricado en los Estados Unidos de América
Primera impresión: mayo de 2000

Los resultados, interpretaciones y conclusiones expresados en este estudio son de la exclusiva responsabilidad de su autor o autores y de ninguna manera pueden ser atribuidos al Banco Mundial, sus organizaciones afiliadas o miembros de su Directorio Ejecutivo o los países que ellos representan. El Banco Mundial no garantiza la exactitud de los datos incluidos en esta publicación y no acepta responsabilidad alguna por las consecuencias de su uso.

El material de esta publicación está registrado como propiedad literaria. El Banco Mundial estimula la divulgación de su trabajo y generalmente autorizará de inmediato la reproducción de partes de éste.

El Banco Mundial otorga autorización para *fotocopiar* artículos para uso interno o personal, para el uso interno o personal de clientes específicos o para uso educacional en clases, siempre que se pague la tarifa correspondiente directamente a Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA.; teléfono 978-750-8400, fax 978-750-4470. Sírvase comunicarse con el Copyright Clearance Center antes de fotocopiar artículos.

Para obtener autorización para *reimprimir* artículos o capítulos, sírvase enviar un fax con la solicitud y la información completa al Reproduction Department, Copyright Clearance Center, fax 978-750-4470.

Todas las demás consultas sobre derechos y licencias deben dirigirse a Office of the Publisher, Banco Mundial, a la dirección recién indicada, o por fax al 202-522-2422

ISBN 0-8213-4697-0

Datos del catálogo de publicación de la Biblioteca del Congreso

Baker, Judy L., JBaker2@worldbank.org

Evaluating the impact of development projects on poverty : a handbook
for practitioners / Judy L. Baker

p. cm. — (Instrucciones en preparación)

Incluye referencias bibliográficas.

ISBN 0-8213-4697-0

1. Economic development projects—Evaluation—Handbooks, manuals, etc.

2. Poor—Developing countries. I. Title II. Instrucciones en preparación
(Washington, D.C.)

HD75.9 .B35 2000
338.9'0068'4—dc21

00—028325

Índice

Prólogo	vi
Agradecimientos	viii
1. Definición de conceptos y técnicas para la evaluación de impacto	1
2. Pasos clave en el diseño e implementación de una evaluación de impacto	15
3. Aplicación de métodos analíticos para la evaluación de impacto: Estudio de caso	38
4. Cómo aprovechar las evaluaciones de impacto de “práctica óptima”	62
Bibliografía	79
Anexos	
Anexo 1: Estudios de casos	89
1.1 Evaluación de los beneficios de los programas de trabajo público con salario mínimo para los pobres: Programa TRABAJAR de Argentina	89
1.2 ¿Las microfinanzas ayudan realmente a los pobres? Nuevas evidencias de programas representativos en Bangladesh	96
1.3 Alimentos por educación en Bangladesh: Evaluación de un programa social focalizado cuando la asignación está descentralizada	100
1.4 Evaluación del Fondo de Inversión Social de Bolivia	104
1.5 El impacto de los programas activos de trabajo: República Checa	108
1.6 Impacto del crédito con educación en la nutrición de las madres y sus hijos menores: Programa del Banco Rural Lower Pra en Ghana	114
1.7 Libros de textos y puntuaciones de pruebas: Evidencia de una evaluación prospectiva en Kenya	118
1.8 Evaluación del proyecto de extensión agrícola de Kenya	123
1.9 El impacto del programa de readiestramiento de México (PROBECAT) en el empleo y los salarios	129
1.10 México, programa nacional de educación, salud y nutrición (PROGRESA)	134
1.11 Evaluación de la reforma escolar de Nicaragua: Un planteamiento cuantitativo y cualitativo combinado	138
1.12 Mejoramiento de la instrucción primaria en matemáticas de Nicaragua: Un estudio experimental del impacto de los libros de texto y la radio en el rendimiento	144
1.13 El impacto de sistemas de recuperación de	

costos alternativos en el acceso y equidad en Níger	149
1.14 Resultados de la instrucción en escuelas primarias filipinas: Evaluación del impacto de cuatro experimentos	153
1.15 Evaluación del impacto de los proyectos de caminos rurales en la pobreza en Vietnam	158
Anexo 2: Muestra de términos de referencia	162
2.1 Proyecto de Nutrición y Desarrollo de la Primera Infancia de Uganda	162
2.2. Evaluación del impacto de los caminos rurales: Datos básicos de 1997 para Vietnam	180
Anexo 3: Muestra de presupuesto de la evaluación del impacto de un programa de alimentación escolar	186
Anexo 4: Indicadores de impacto: Evaluación del Fondo de Inversión Social de Bolivia	188
Anexo 5: Plantilla del marco lógico para el diseño de un proyecto Resumen del Documento de realización del proyecto o el Documento de evaluación del proyecto	193
Anexo 6: Matriz de análisis	196
Recuadros	
1.1. El problema del sesgo de selección	5
1.2. Resumen de métodos cuantitativos para evaluar el impacto de un programa	6
1.3. Resumen de los métodos usados para evaluar políticas de ajuste	10
2.1 Principales pasos en el diseño e implementación de una evaluación de impacto	16
2.2 Puntos clave para identificar las fuentes de datos disponibles para la evaluación de impacto	20
3.1. Pasos en la correspondencia de puntuación de la propensión	47
3.2 Fuentes de sesgo en estimaciones simplistas del impacto de PROSCOL	50
3.3. Ejecutando una doble diferencia	53
3.4 Mediciones de pobreza	56
3.5 Comparación de la pobreza con el programa y sin él	57
Cuadros	
2.1 Métodos de evaluación y las correspondientes	

necesidades de datos	26
2.2 Principales instrumentos de recopilación de datos para la evaluación de impacto	31
4.1 Resumen de evaluaciones de impacto de "práctica óptima"	64
4.2 Resumen de costos estimados de diversas evaluaciones de impacto del Banco Mundial	75

Prólogo

A pesar de los miles de millones de dólares que se emplean en ayuda para el desarrollo cada año, aún se conoce muy poco acerca del efecto real de los proyectos en los pobres. Hay evidencias generalizadas sobre los beneficios que el crecimiento económico, las inversiones en capital humano y el suministro de redes de seguridad tienen en los pobres. Pero para un programa o proyecto específico en un país determinado, ¿la intervención está produciendo los beneficios previstos y cuál fue el efecto general en la población? ¿Se podría diseñar mejor el programa o proyecto para lograr los resultados previstos? ¿Se están empleando los recursos en forma eficiente? Estos son los tipos de preguntas que sólo se pueden responder mediante una evaluación de impacto, un enfoque que mide los resultados de la intervención de un programa aisándolo de otros posibles factores.

Muchos gobiernos, instituciones y administradores de proyectos se muestran reticentes a realizar evaluaciones de impacto, porque las consideran costosas, prolongadas y técnicamente complejas y porque los resultados pueden ser políticamente delicados, especialmente si son negativos. Muchas evaluaciones también reciben críticas porque los resultados se obtienen demasiado tarde, no responden las preguntas adecuadas o no se realizaron con suficiente rigor analítico. Otra restricción frecuente es la limitada disponibilidad y calidad de los datos.

Sin embargo, con una planificación adecuada y oportuna, el apoyo de las autoridades responsables y una inversión relativamente pequeña en comparación con el costo total del proyecto, una evaluación rigurosa puede constituirse en una herramienta muy poderosa para evaluar la conveniencia y eficacia de los programas. La evaluación de impacto es especialmente importante en los países en desarrollo, donde los recursos son escasos y cada dólar gastado debe maximizar su efecto en la reducción de la pobreza. Si los programas están mal diseñados, no llegan a los beneficiarios previstos o despilfarran los recursos, con la información adecuada es posible rediseñarlos, mejorarlos o eliminarlos si se estima necesario. El conocimiento obtenido de los estudios para evaluar los efectos también proporcionará información decisiva para el diseño adecuado de programas y proyectos futuros.

Este manual intenta proporcionar a los administradores de proyectos y a los analistas de políticas las herramientas necesarias para evaluar el impacto de los proyectos. Está destinado a lectores con conocimientos generales de estadísticas. Para algunos de los métodos estadísticos más profundos analizados, el lector es remitido a la literatura técnica sobre el tema. En el Capítulo 1 se presenta una visión general de los conceptos y métodos, en el Capítulo 2 se analizan los pasos clave y temas relacionados que hay que considerar durante la implementación, en el Capítulo 3 se ilustran diversas técnicas analíticas a través de un estudio de casos y en el Capítulo 4 se incluye un análisis de las lecciones aprendidas de un valioso conjunto de evaluaciones de “prácticas óptimas” de los proyectos de pobreza que se revisaron para este manual. Los estudios de casos, incluidos en el Anexo I, se seleccionaron de una variedad de evaluaciones realizadas por el Banco Mundial, otras agencias de donantes, instituciones de investigación y empresas consultoras privadas. Se eligieron por su rigor metodológico, en un intento por cubrir una amplia combinación de escenarios de países, tipos de proyectos y metodologías de evaluación. En los Anexos

también se incluyen muestras de los principales componentes que serían necesarios al planificar cualquier evaluación de impacto: muestras de términos de referencia, un presupuesto, indicadores de impacto, un marco lógico y una matriz de análisis.

Aunque las técnicas usadas en la evaluación del impacto son similares en todos los sectores y subgrupos de la población, las ilustraciones de las metodologías y los ejemplos de casos que se muestran en el manual se centran en la evaluación del efecto de los proyectos orientados hacia los pobres. El efecto en la pobreza puede incluir una amplia gama de proyectos y preguntas sobre evaluación, como la medición del efecto de los programas de microfinanzas en el ingreso de los hogares, la repercusión de un programa de capacitación en el empleo, el impacto de un programa de alimentación escolar en la asistencia de los alumnos o las repercusiones de la construcción de caminos rurales en el bienestar de los hogares.

Sin importar el tipo de proyecto ni las preguntas abordadas, el diseño de cada evaluación del impacto será único y dependerá de factores como el tipo de datos disponibles, la capacidad local y aspectos presupuestarios y de secuencia cronológica. Por último, las evaluaciones que generen resultados de alta calidad, creíbles y generalizables para las autoridades responsables requerirán un fuerte apoyo financiero y político, una planificación oportuna y cuidadosa, la participación de las partes interesadas en el diseño de los objetivos y enfoque del estudio, datos adecuados, una combinación apropiada de metodologías (incluidas técnicas cuantitativas y cualitativas), la aplicación rigurosa de estas técnicas y una buena comunicación entre los miembros del equipo durante el proceso.

Agradecimientos

La preparación de este libro se benefició de las valiosas contribuciones de un equipo básico. Deseo agradecer a los siguientes miembros del equipo por sus valiosos aportes escritos y útiles comentarios durante el proceso: Gillette Hall (estudios de casos, lecciones aprendidas), Julia Lane (estudios de casos, lecciones aprendidas), Martin Ravallion (estudio de casos de métodos analíticos) y Laura Rawlings (aspectos de implementación, lecciones aprendidas); y el trabajo sobre evaluación del impacto realizado por Kene Ezemenari, Gloria Rubio, Anders Rudqvist y K. Subbarao. La investigación básica fue realizada por Matthew Fleming y Samir Stewart. El libro contó con el respaldo conjunto del departamento de la Región de América Latina y el Caribe y la Red de Reducción de la Pobreza y Administración Económica del Banco Mundial bajo la dirección de Norman Hicks, Guillermo Perry y Michael Walton. El trabajo también se benefició en alto grado de los comentarios recibidos de Omar Arias, Sabina Alkire, Michael Bamberger, Soniya Carvalho, Wendy Cunningham, Norman Hicks, Shahidur Khandker, Norbert Schady y Quentin Wodon.

Capítulo 1

Definición de conceptos y técnicas para la evaluación de impacto

En la literatura, una evaluación global se define como una evaluación que incluye la supervisión, evaluación de los procesos, evaluación de costos-beneficios y evaluación de impacto. No obstante, cada uno de estos componentes es completamente diferente. La supervisión ayudará a evaluar si un programa se está implementando de acuerdo a lo planificado. Un sistema de supervisión de programas permite una retroalimentación constante sobre el estado en que se encuentra la implementación del programa e identifica los problemas específicos a medida que surgen. La evaluación de los procesos se relaciona con la forma en que funciona el programa y se centra en los problemas de la entrega de servicios. En las evaluaciones de costos-beneficios o eficacia en función de los costos se estiman los costos de los programas (monetarios o no monetarios), en particular su relación con respecto a usos alternativos de los mismos recursos y a los beneficios que produce el programa. Por último, la evaluación del impacto tiene el objeto de determinar en forma más general si el programa produjo los efectos deseados en las personas, hogares e instituciones y si esos efectos son atribuibles a la intervención del programa. Las evaluaciones de impacto también permiten examinar consecuencias no previstas en los beneficiarios, ya sean positivas o negativas. Para este manual, es de particular interés la medida en que los beneficios del proyecto llegan a los pobres y el efecto de estos beneficios en su bienestar. Algunas de las preguntas que se abordan en la evaluación del impacto incluyen las siguientes: ¿Cómo afectó el proyecto a los beneficiarios? ¿Algún mejoramiento fue el resultado directo del proyecto o se habría producido de todas formas? ¿Se podría modificar el diseño del programa para mejorar sus repercusiones? ¿Se justificaban los costos?

Sin embargo, estas preguntas no se pueden medir simplemente a través del resultado de un proyecto. Puede haber otros factores o sucesos que estén correlacionados con los resultados, sin ser causados por el proyecto. Para asegurar un rigor metodológico, una evaluación del impacto debe estimar el escenario contrafactual o simulado alternativo, es decir, lo que habría ocurrido si el proyecto nunca se hubiera realizado o lo que habría ocurrido normalmente. Por ejemplo, si un recién graduado de un programa de capacitación laboral obtiene empleo, ¿es un resultado directo del programa o habría encontrado empleo de todas formas? Para determinar el escenario contrafactual, es necesario separar el efecto de las intervenciones de otros factores; una tarea algo compleja. Esto se logra con la ayuda de grupos de comparación o de control (aquellos que no participan en un programa ni reciben beneficios), que luego se comparan con el grupo de tratamiento (personas que reciben la intervención). Los grupos de control se seleccionan en forma aleatoria de la misma población que los participantes del programa, mientras que el grupo de comparación es simplemente el grupo que no recibe el programa que se está investigando. Los grupos de comparación y de control deben ser semejantes al grupo de tratamiento en todo aspecto y la única diferencia entre los grupos es la participación en el programa.

Determinar el escenario contrafactual es esencial para el diseño de la evaluación. Esto se puede realizar usando diversas metodologías que entran en dos categorías generales: diseños experimentales (aleatorios) y diseños cuasi experimentales (no aleatorios). Sin embargo, es bastante complicado separar el efecto del programa de las condiciones hipotéticas que pueden verse afectadas por la historia, el sesgo de selección y la contaminación. También se pueden usar métodos cualitativos y participativos para evaluar el impacto. Estas técnicas con frecuencia proporcionan información decisiva sobre las perspectivas de los beneficiarios, el valor que los programas revisten para éstos, los procesos que pueden haber afectado los resultados y una interpretación más profunda de los resultados observados en el análisis cuantitativo. A continuación se analizan con más detalle las fortalezas y debilidades de cada uno de estos métodos. Como lo descubrirá el lector, ninguna técnica es perfecta y, por lo tanto, el evaluador debe tomar decisiones frente a las ventajas y desventajas de cada uno de los métodos elegidos. Sin embargo, una planificación oportuna y cuidadosa proporcionará muchas más opciones metodológicas para diseñar la evaluación.

Diseños experimentales

En general se considera que los diseños experimentales, conocidos también como aleatorización, son las metodologías de evaluación más sólidas. Al distribuir aleatoriamente la intervención entre los beneficiarios calificados, el proceso de asignación mismo crea grupos de tratamiento y de control comparables que son estadísticamente equivalentes entre sí, a condición de que las muestras sean de tamaño adecuado. Se trata de un resultado muy convincente porque, en teoría, los grupos de control generados mediante asignación aleatoria sirven como un escenario contrafactual perfecto, sin los dificultosos problemas de sesgo de selección que existen en todas las evaluaciones. La principal ventaja de esta técnica es la simplicidad en la interpretación de los resultados, puesto que el efecto del programa sobre el resultado que se evalúa se puede medir a través de la diferencia entre las medias de las muestras del grupo de tratamiento y el grupo de control. Un ejemplo es la evaluación de libros de texto de Kenya: los evaluadores seleccionaron una distribución aleatoria de lugares donde aplicar el programa, realizaron una encuesta básica o de referencia, crearon grupos de control y luego administraron el tratamiento, que en este caso fue la entrega de libros de texto. Los grupos de control y de tratamiento permitieron que los evaluadores determinaran claramente el efecto de los libros de texto en el aprendizaje de los alumnos.

Aunque los diseños experimentales se consideran el método óptimo para estimar el impacto de un proyecto, en la práctica conllevan varios problemas. Primero, la aleatorización podría ser poco ética debido a la negación de beneficios o servicios a miembros de la población de por sí calificados para el estudio. Como un ejemplo extremo se podría citar la negación de tratamiento médico que podría salvar la vida de algunos miembros de la población. Segundo, puede ser políticamente difícil proporcionar una intervención a un grupo y no a otro. Tercero, el alcance del programa podría significar que no hubiera grupos sin tratamiento, como en el caso de un proyecto o cambio de política de amplio alcance. Entre los ejemplos se incluye un préstamo de ajuste o programas aplicados a nivel nacional.

Cuarto, durante el experimento los individuos de los grupos de control podrían cambiar ciertas características que los identifican, lo que podría invalidar o contaminar los resultados. Por ejemplo, si las personas se trasladan hacia un área de proyecto o salen de él, podrían ingresar o salir del grupo de tratamiento o de control. Alternativamente, las personas a las que se les niega el beneficio del programa pueden buscarlo a través de otras fuentes, o bien aquellas a las que se les ofrece un programa podrían no aceptar la intervención. Quinto, podría resultar difícil garantizar que la asignación sea realmente aleatoria. Un ejemplo de esto podrían constituirlo los administradores que excluyen solicitantes de alto riesgo para lograr mejores resultados. Y por último, los diseños experimentales pueden ser costosos y prolongados en ciertas situaciones, especialmente en la recopilación de nuevos datos.

Con una planificación cuidadosa, se pueden abordar algunos de estos problemas al momento de implementar los diseños experimentales. Una forma es la selección aleatoria de los beneficiarios. Esto se puede aplicar para proporcionar un mecanismo de distribución políticamente transparente y la base de un diseño de evaluación sólido, puesto que las restricciones de presupuesto o de información con frecuencia hacen imposible identificar y llegar de manera precisa a los beneficiarios más calificados. Una segunda forma es ingresar los grupos de control al programa en una etapa posterior, una vez que se ha diseñado y se ha iniciado la evaluación. Con esta técnica, la selección aleatoria determina *cuándo* el beneficiario calificado recibe el programa y no *si* lo recibe. Esto se aplicó en la evaluación de un programa de nutrición en Colombia, proporcionando la ventaja adicional de abordar preguntas con respecto al tiempo necesario para que el programa sea eficaz para reducir la desnutrición (McKay, 1978). Por último, se puede aplicar la aleatorización dentro de un subconjunto de beneficiarios igualmente calificados, llegando al mismo tiempo a todos los más calificados y negando los beneficios a los menos calificados, como se hizo con los proyectos de educación en la región del Chaco para la evaluación del fondo social en Bolivia (Pradhan, Rawlings y Ridder, 1998). Sin embargo, si se implementa esta última sugerencia, se debe tener en cuenta que los resultados que arroje la evaluación serán válidos para el grupo del cual se tomó la muestra generada aleatoriamente.

Diseños cuasi experimentales

Se pueden emplear métodos cuasi experimentales (no aleatorios) para realizar una evaluación cuando es imposible crear grupos de tratamiento y de comparación a través de un diseño experimental. Estas técnicas generan grupos de comparación que se asemejan al grupo de tratamiento, al menos en las características observadas, usando metodologías econométricas que incluyen métodos de pareo, métodos de doble diferencia, métodos de variables instrumentales y comparaciones reflexivas (véase el Recuadro 1.2). Cuando se usan estas técnicas, los grupos de tratamiento y de comparación por lo general se seleccionan *después* de la intervención usando métodos no aleatorios. Por lo tanto, se deben aplicar controles estadísticos para abordar las diferencias entre los grupos de tratamiento y de comparación y emplear técnicas de pareo sofisticadas para crear un grupo de comparación que sea lo más similar posible al grupo de tratamiento. En algunos casos también se selecciona un grupo de comparación antes del tratamiento, aunque la selección no es aleatoria.

La ventaja principal de los diseños cuasi experimentales es que se pueden basar en fuentes de datos existentes y, por lo tanto, a menudo son más rápido y menos costosos en implementar. Además, se pueden realizar una vez que el programa se ha implementado, a condición de que existan suficientes datos. Las desventajas principales de las técnicas cuasi experimentales son que (a) con frecuencia se reduce la confiabilidad de los resultados, puesto que la metodología es menos sólida estadísticamente, (b) los métodos pueden ser estadísticamente complejos y (c) conllevan un problema de sesgo de selección. Al generar un grupo de comparación en lugar de asignarlo aleatoriamente, hay muchos factores que pueden afectar la confiabilidad de los resultados. La complejidad estadística requiere conocimientos especializados considerables en el diseño de la evaluación y el análisis e interpretación de los resultados. Esto no siempre es posible, especialmente en las condiciones de los países en desarrollo.

El tercer problema de sesgo se relaciona con la medida en la cual los subgrupos de una población beneficiaria participan en forma diferenciada en el programa, afectando así la muestra y, finalmente, los resultados. Hay dos tipos de sesgo: aquellos causados por las diferencias en los elementos observables o algún elemento de los datos y aquellos causados por las diferencias en los elementos no observables (no en los datos), lo que con frecuencia se denomina sesgo de selección (Recuadro 1.1). Un sesgo observable podría incluir los criterios de selección mediante los cuales se escoge a un individuo, como ubicación geográfica, asistencia a la escuela o participación en el mercado laboral. Los no observables que pueden sesgar los resultados de un programa podrían incluir la capacidad individual, la disposición al trabajo, los vínculos familiares y un proceso subjetivo (con frecuencia guiado por la política) de seleccionar individuos para un programa. Ambos tipos de sesgos pueden generar resultados inexactos, como subestimar y sobrestimar los efectos reales del programa, efectos negativos cuando los efectos reales del programa son positivos (y viceversa) y efectos estadísticamente insignificante cuando los efectos reales del programa son relevantes y viceversa. (Véase, por ejemplo, LaLonde, 1986, Fraker y Maynard, 1987, LaLonde y Maynard, 1987, y Friedlander y Robins, 1995). Es posible controlar por el sesgo mediante técnicas estadísticas, como variables de comparación e instrumentales, pero es muy difícil eliminarlo por completo, pero sigue siendo un importante desafío para los investigadores en el campo del análisis de los efectos.

Entre las técnicas de diseño cuasi experimental en general se considera que las técnicas de comparación pareada son la alternativa subóptima al diseño experimental. Gran parte de la literatura sobre metodologías de evaluación se centra en el uso de este tipo de evaluaciones, lo que indica el frecuente uso de las comparaciones pareadas y los numerosos desafíos que plantea el contar con grupos de comparación poco adecuados. En los últimos años se han producido significativos avances en las técnicas de correspondencia de puntuación de la propensión (Rosenbaum y Rubin, 1985; Jalan y Ravallion, 1998). Este método es muy atractivo para los evaluadores que tienen restricciones de tiempo y no tienen la ventaja de contar con datos básicos (de referencia), dado que se pueden usar con una simple sección transversal de datos. Sin embargo, para esta técnica se debe contar con los datos adecuados, porque se basa en tomar nuevas muestras de los beneficiarios durante la aplicación en terreno de una encuesta más amplia, “haciéndolos corresponder” luego con un grupo de

comparación seleccionada de la muestra básica más amplia de la iniciativa global, frecuentemente una encuesta domiciliaria a nivel nacional. Dado el crecimiento de las aplicaciones de grandes encuestas en los países en desarrollo, como las Encuestas Sobre las

Recuadro 1.1 El problema del sesgo de selección

El sesgo de selección se relaciona con los elementos no observables que pueden sesgar los resultados (por ejemplo, capacidad individual, condiciones preexistentes). Los experimentos aleatorizados resuelven el problema del sesgo de selección al generar un grupo de control experimental de personas que habría participado en un programa, pero al que aleatoriamente se le negó el acceso al programa o tratamiento. La asignación aleatoria no elimina el sesgo de selección, sino que equilibra el sesgo entre las muestras de participantes y no participantes. En los diseños cuasi experimentales, los modelos estadísticos (por ejemplo, de pareo, dobles diferencias, variables instrumentales) se acercan a esto modelando los procesos de selección para llegar a una estimación insesgada usando datos no experimentales. La idea general es comparar los participantes y no participantes del programa, manteniendo constantes los procesos de selección. La validez de este modelo depende de qué tan bien se especifique el modelo.

Un buen ejemplo es el salario de las mujeres. Los datos representan mujeres que optan por trabajar. Si se tomara esta decisión, podríamos ignorar el hecho de que no se observan todos los salarios y usar la regresión ordinaria para estimar un modelo salarial. Sin embargo, la decisión de trabajar por parte de las mujeres no se toma aleatoriamente; es poco probable que las mujeres que tuvieran bajos salarios optaran por trabajar, porque su salario de reserva personal es mayor que el salario ofrecido por los empleadores. Así, la muestra de los salarios observados para las mujeres tendría un sesgo ascendente.

Esto se puede corregir si hay algunas variables que afecten en forma significativa las posibilidades de observación (el salario de reserva), pero no el resultado que se está estudiando (el salario ofrecido). Esta variable podría ser el número de hijos en casa.

Fuente: Greene (1997)

Condiciones de Vida de propósitos múltiples, este método de evaluación parece particularmente prometedor. Un buen ejemplo es la evaluación de un programa de obras públicas, TRABAJAR, en Argentina (Jalan y Ravallion, 1998, Anexo 1.1 y Capítulo 4).

Recuadro 1.2 Resumen de métodos cuantitativos para evaluar el impacto de un programa

A continuación se analizan los principales métodos de evaluación de impacto. Dado que ningún método es perfecto, siempre es conveniente hacer una triangulación.

Diseños de control experimental o aleatorio

- *Aleatorización*, en la cual la selección para los grupos de tratamiento y de control es aleatoria dentro de algún conjunto bien definido de personas. En este caso, no debería haber diferencia (en el valor esperado) entre los dos grupos, aparte del hecho de que el grupo de tratamiento tuvo acceso al programa (aunque puede haber diferencias debido a un error de muestreo; mientras mayor sea el tamaño de las muestras de tratamiento y control, menor será el error).

Diseños no experimentales o cuasi experimentales

- *Métodos de pareo o controles contruidos*, en los cuales se intenta obtener una comparación ideal que corresponda al grupo de tratamiento de una encuesta más amplia. El tipo de correspondencia de uso más generalizado es la *correspondencia de puntuación de la propensión*, en la cual el grupo de comparación se compara con el grupo de tratamiento sobre la base de un conjunto de características observadas o bien usando la “puntuación de la propensión” (probabilidad proyectada de participar dadas las características observadas); mientras más precisa sea la puntuación de la propensión, mejor será la correspondencia. Un buen grupo de comparación proviene del mismo entorno económico y se le ha aplicado el mismo cuestionario por parte de entrevistadores similarmente capacitados que el grupo de tratamiento.
- *Métodos de doble diferencia o diferencia en las diferencias*, en los cuales se compara un grupo de tratamiento y uno de comparación antes (primera diferencia) y después de un programa (segunda diferencia). Se deben eliminar los comparadores cuando se utilizan puntuaciones de la propensión y si tienen puntuaciones fuera del margen observado para el grupo de tratamiento.
- *Métodos de variables instrumentales o control estadístico*, en los cuales se usa una o más variables que influyen en la participación, pero no en los resultados dada la participación. Esto identifica la variación exógena en los resultados atribuibles al programa, reconociendo que su establecimiento no es aleatorio sino intencional. Las “variables instrumentales” se usan primero para predecir la participación en el programa y luego se observa cómo varía el indicador de resultados con los valores proyectados.
- *Comparaciones reflexivas*, en las cuales se realiza una encuesta básica o de referencia de los participantes antes de la intervención y luego se realiza una encuesta de seguimiento. La encuesta básica proporciona el grupo de comparación y el efecto se mide mediante el cambio en los indicadores de resultado antes y después de la intervención.

Métodos cualitativos

Para realizar una evaluación del impacto también se usan técnicas cualitativas, en un intento por determinar el efecto basándose en algo diferente al escenario contrafactual para realizar una inferencia causal (Mohr, 1995). En su lugar, se trata de comprender los procesos, comportamientos y condiciones como las perciben los individuos o grupos estudiados (Valadez y Bamberger, 1994). Por ejemplo, los métodos cualitativos y, en particular, la observación de los participantes, puede proporcionar información sobre las formas en que los hogares y las comunidades locales perciben un proyecto y cómo se ven afectados por éste. Puesto que medir el escenario contrafactual es esencial para las técnicas de análisis de los efectos, los diseños cualitativos en general se han usado en conjunto con otras técnicas de evaluación. El enfoque cualitativo utiliza métodos relativamente flexibles durante el diseño, recopilación de datos y análisis. Los datos cualitativos también se pueden cuantificar. Entre las metodologías que se usan en las evaluaciones cualitativas de los efectos se encuentran las técnicas elaboradas para la evaluación rural rápida, las que se basan en el conocimiento de los participantes sobre las condiciones que rodean al proyecto o el programa que se está evaluando, o las evaluaciones participativas, en que las partes interesadas intervienen en todas las etapas de la evaluación, determinando los objetivos del estudio, identificando y seleccionando los indicadores que se usarán y participando en la recopilación y análisis de datos. Para obtener un análisis detallado sobre los métodos participativos, véase Banco Mundial (1996), *The World Bank Participation Sourcebook*.

Las ventajas de las evaluaciones cualitativas son su flexibilidad y la posibilidad de ser adaptadas específicamente a las necesidades de la evaluación usando enfoques adaptables, de realizarlas usando técnicas inmediatas y de mejorar en forma significativa los resultados de una evaluación del impacto al proporcionar una mayor comprensión de las percepciones y prioridades de las partes interesadas y las condiciones y procesos que pueden haber afectado las repercusiones de un programa.

Entre las principales desventajas se encuentra la subjetividad involucrada en la recopilación de datos, la falta de un grupo de comparación y la falta de solidez estadística dados los tamaños de las muestra en su mayoría reducidos, lo que hace difícil generalizar los resultados para una población representativa y más numerosa. La validez y confiabilidad de los datos cualitativos depende en gran medida de la habilidad metodológica, sensibilidad y capacitación del evaluador. Si el personal en terreno no es sensible a las normas y prácticas sociales y culturales específicas y a los mensajes no verbales, los datos reunidos pueden ser mal interpretados. Y por último, sin un grupo de comparación, es imposible determinar el escenario contrafactual y, por lo tanto, la causalidad del efecto del proyecto.

Integración de métodos cuantitativos y cualitativos

Aunque existe abundante literatura en que se comparan los métodos cuantitativos con los cualitativos en la evaluación del impacto, cada vez hay más aceptación de que es necesario integrar los dos enfoques. Las evaluaciones de impacto que se basan en datos cuantitativos de muestras estadísticamente representativas son más adecuadas para evaluar la causalidad usando métodos econométricos o llegando a conclusiones que se pueden generalizar. Sin

embargo, los métodos cualitativos permiten estudiar cabalmente los temas, casos o hechos seleccionados y pueden proporcionar información decisiva sobre las perspectivas de los beneficiarios, la dinámica de una determinada reforma o los motivos de ciertos resultados observados en un análisis cuantitativo. Existen significativas ventajas y desventajas en la selección de una técnica en lugar de otra.

La integración de las evaluaciones cuantitativas y cualitativas con frecuencia puede ser el mejor vehículo para satisfacer las necesidades de un proyecto. Al combinar los dos enfoques, los métodos cualitativos se pueden usar para informar las preguntas clave sobre la evaluación del impacto, examinar el cuestionario o la estratificación de la muestra cuantitativa y analizar el marco social, económico y político dentro del cual se lleva a cabo un proyecto. Los métodos cuantitativos, en tanto, se pueden usar para informar las estrategias de recopilación de datos cualitativos, diseñar la muestra para informar la medida en que los resultados observados en el trabajo cualitativo son aplicables a una población de mayor tamaño al usar una muestra estadísticamente representativa. Y finalmente, el análisis estadístico se puede usar para controlar por las características de los hogares y las condiciones socioeconómicas de diferentes áreas de estudio, con lo que se eliminan las explicaciones alternativas de los resultados observados.

Existen varias ventajas en usar enfoques integrados en la investigación que se analizan en Bamberger (2000), los que también se aplican a las evaluaciones de impacto.

Entre ellas:

- Se pueden incorporar controles de coherencia con la ayuda de procedimientos de triangulación que permiten realizar dos o más estimaciones independientes de las variables clave (como ingreso, opiniones sobre los proyectos, motivos para usar o no los servicios públicos y efecto específico de un proyecto).
- Se pueden obtener diferentes perspectivas. Por ejemplo, aunque los investigadores podrían considerar el ingreso o el consumo como indicadores clave del bienestar de un hogar, los estudios de casos podrían revelar que las mujeres están más preocupadas por la vulnerabilidad (definida como la falta de acceso a sistemas de apoyo social en tiempos de crisis), impotencia o exposición a la violencia.
- Los análisis se pueden conducir en diferentes niveles. Los métodos de encuestas pueden proporcionar buenas estimaciones del bienestar individual, familiar y a nivel de la comunidad, pero son mucho menos eficaces para analizar los procesos sociales (conflicto social, motivos para usar o no usar los servicios, etc.) o para el análisis institucional (qué tan eficazmente funcionan los servicios de salud, educación, crédito y otros y cómo son percibidos por la comunidad). Existen muchos métodos cualitativos que fueron diseñados para analizar temas como el proceso social, comportamiento institucional, estructura social y conflictos.
- Se pueden proporcionar oportunidades de retroalimentación para interpretar los resultados. Los informes de encuestas a menudo incluyen referencias a aparentes

incoherencias en los resultados o a diferencias interesantes entre las comunidades o grupos que no se pueden explicar con los datos. En la mayoría de las investigaciones cuantitativas, una vez que finaliza la fase de recolección de datos no es posible regresar al terreno para verificar esos aspectos. La mayor flexibilidad de la investigación cualitativa significa que a menudo es posible regresar al terreno para reunir datos adicionales. Los investigadores de encuestas también utilizan métodos cualitativos para verificar los valores atípicos: respuestas que se desvían de los patrones generales. En muchos casos, el analista de datos debe tomar una decisión arbitraria sobre si se debe excluir o no a un hogar o comunidad que muestra condiciones significativamente sobre o bajo la norma (bajo el supuesto de que refleja un error de información) o si se deben ajustar las cifras. Los métodos cualitativos permiten un seguimiento rápido en terreno para verificar estos casos.

En la práctica, la integración de los métodos cuantitativos y cualitativos se debe realizar durante cada paso de la evaluación del impacto. En el Capítulo 2 se mencionan muchas oportunidades para hacerlo. Como ilustración, el Caso de Reforma de la Autonomía Escolar de Nicaragua proporciona un buen ejemplo de métodos integrados. Se emplearon métodos cualitativos con un diseño cuasi experimental para determinar la relación entre una administración descentralizada y el aprendizaje y para generalizar los resultados para distintos tipos de escuelas. Además, se utilizaron técnicas cualitativas, incluida una serie de decisivas entrevistas a informantes y conversaciones con grupos representativos a nivel de diferentes miembros del personal de las escuelas y padres para analizar el contexto en que se introdujo la reforma, examinar la dinámica de toma de decisiones en cada escuela y evaluar las perspectivas de diferentes participantes de la comunidad escolar sobre el proceso de autonomía (véase el Anexo 1.11).

Otros enfoques para la evaluación de impacto

Hay otros dos temas que son especialmente pertinentes al momento de analizar la evaluación de las repercusiones de los proyectos en la pobreza. (a) los enfoques para medir el efecto de los programas de ajuste estructural y (b) evaluaciones basadas en la teoría. En ambos se incorporan muchas de las metodologías antes analizadas, pero cada uno utiliza un enfoque diferente.

Evaluación de los programas de ajuste estructural. Se ha discutido mucho sobre el efecto de los programas de ajuste estructural en la pobreza. No obstante, gran parte de la evidencia usada para sostener este debate se basa en supuestos y métodos deficientes. Al igual que con otros proyectos, los cambios de política bajo proyectos de ajuste estructural (a) se deben comparar con casos hipotéticos pertinentes que responderían a la mismas restricciones macroeconómicas y (b) se deben analizar en el contexto de la estructura económica local y se deben basar en la información empírica de encuestas domiciliarias. Sin embargo, esto es muy difícil por tres motivos. En primer lugar, los cambios de políticas podrían tener un efecto a nivel de toda la economía, haciendo imposible encontrar grupos de comparación que no se hayan visto afectados. En segundo lugar, debido a factores exógenos, retrasos, retroalimentaciones y sustituciones, cualquier cambio en el bienestar de los pobres debe interpretarse con extrema prudencia. Y en tercer lugar, es difícil predecir lo

que habría sucedido si no se hubiera realizado el ajuste, esto es, qué políticas alternativas habría seguido un gobierno y cómo habría afectado a los pobres.

En la literatura se han usado diversos enfoques, cada uno con sus propias falencias. En muchos casos, las técnicas son muy similares a las descritas en el Recuadro 1.2. Sin embargo, como se muestra en el Recuadro 1.3, para estimar el escenario contrafactual hay que basarse en amplios supuestos que pueden afectar considerablemente la validez de los resultados. La forma más viable de manejarlo es aislar cambios de políticas específicos que afectarían a la población, como las políticas cambiarias, políticas comerciales, reducciones en el gasto público y reducciones en el empleo del sector público. Sin embargo, incluso con este enfoque puede ser difícil aislar el efecto de políticas específicas. Ejemplos de esto se encuentran en Killick (1995), Poppele, Summarto y Pritchett (1999), Bourguignon, de Melo y Suwa (1991) y Sahn, Dorosh y Younger (1996).

Recuadro 1.3 Resumen de los métodos usados para evaluar políticas de ajuste

Enfoques sin escenario contrafactual

- Los estudios cualitativos que evalúan las condiciones de la población (a menudo mediante la identificación de subgrupos vulnerables) antes, durante y después de las políticas de ajuste se implementan a través de grupos representativos, entrevistas y otras técnicas cualitativas.
- Enfoques “antes y después”, en que se compara el comportamiento de variables clave durante y después de un programa con aquel anterior al programa. En este caso se emplean métodos estadísticos para evaluar si existe algún cambio significativo en algunas variables esenciales en el tiempo. Este enfoque con frecuencia entrega resultados sesgados, porque se supone que de no mediar el programa, los indicadores de comportamiento habrían retomado sus valores del período anterior a la crisis.

Enfoques que generan un escenario contrafactual usando múltiples supuestos

- Modelos computarizados de equilibrio general (CGE) que intentan contrastar los resultados en los grupos de tratamiento y de comparación mediante simulaciones. Estos modelos pretenden rastrear el funcionamiento de la economía real y en general se basan en matrices de contabilidad social detalladas que se recopilan de datos sobre cuentas nacionales, encuestas de gastos familiares y otros datos de encuestas. Los modelos CGE en realidad generan resultados para el escenario contrafactual, aunque la solidez del modelo depende por completo de la validez de los supuestos. Esto puede ser problemático, puesto que las bases de datos con frecuencia son incompletas y muchos de los parámetros no se han estimado con métodos econométricos formales. Además, la generación de los modelos CGE requiere mucho tiempo, es engorrosa y costosa.

- Con y sin comparaciones. Enfoque en que se compara el comportamiento de variables clave de una muestra de países del programa con su comportamiento en países no pertenecientes al programa (un grupo de comparación). Esta es una forma de abordar la pregunta del escenario contrafactual usando las experiencias del grupo de comparación como un sustituto de lo que hubiera sucedido de lo contrario en los países del programa. Sin embargo, es bastante difícil obtener un verdadero grupo de comparación. El método supone que país del programa sólo se distingue del grupo de comparación por la adopción de un programa de ajuste y que el entorno externo afecta de la misma forma a los dos grupos.
- Controles estadísticos. Constan de regresiones que controlan por las diferencias en las condiciones iniciales y las políticas adoptadas en los países del programa y no pertenecientes a éste. En este método se identifican las diferencias entre los países del programa y no pertenecientes a éste en el período anterior al programa y luego se controlan estadísticamente estas diferencias para aislar los efectos de los programas en el comportamiento posterior a la reforma.

Evaluación basada en la teoría. La premisa de las evaluaciones basadas en la teoría es que los programas y proyectos se basan en teoría explícita o implícita acerca de cómo y por qué funcionará un programa. Entonces, la evaluación se basará en apreciar cada teoría y los supuestos acerca de un programa durante la fase de implementación, en lugar de hacerlo en un momento intermedio o cuando el proyecto ha finalizado. Al diseñar la evaluación, la teoría implícita se presenta en la forma de muchos micropasos, creándose entonces los métodos para que la recolección y el análisis de datos hagan un seguimiento de la exposición de los supuestos. Si los sucesos no resultan como se esperaba, la evaluación puede afirmar con un cierto nivel de confianza dónde, por qué y cómo ocurrió la falla.

El enfoque se centra en las respuestas de las personas a las actividades del programa. Las teorías dirigen la atención del evaluador a los probables tipos de efectos en el corto y más largo plazo. Algunas de las ventajas son, en primer lugar, que la evaluación proporciona indicios oportunos de la eficacia del programa durante la implementación del proyecto. Si hay fallas durante esta fase, es posible solucionarlas durante el proceso. En segundo lugar, el enfoque permite explicar cómo y por qué se produjeron los efectos. Si los sucesos resultan como se esperaba, la evaluación puede indicar con cierta confianza cómo se generaron los efectos. Siguiendo la secuencia de las etapas, es posible hacer un seguimiento de los micropasos que llevan desde los aportes del programa hasta los resultados.

Las falencias del enfoque son similares a muchas de las demás metodologías.

Específicamente, (a) la identificación de los supuestos y teorías puede ser inherentemente complejo; (b) los evaluadores pueden tener problemas para medir cada paso, a menos que cuenten con los instrumentos y datos adecuados, (c) se podrían enfrentar problemas al examinar la iniciativa, porque los planteamientos teóricos podrían ser demasiado generales y estar contruidos en forma demasiado inexacta para permitir un estudio bien definido y (d) podría haber problemas de interpretación que dificultarían la generalización a partir de los resultados (véase Weiss, 1998).

El Departamento de Operaciones y Evaluación del Banco Mundial está probando un ejemplo de la técnica basada en teoría para evaluar el impacto de los fondos de inversión social en los procesos de toma de decisiones a nivel de la comunidad, las estructuras y relaciones tradicionales de poder y la capacidad, confianza y bienestar de la comunidad. Esto se basará en la teoría de que los grupos prioritarios pueden implementar con eficacia un proyecto y manejar y mantener la inversión creada por el proyecto. Se ha establecido un conjunto de sub-supuestos y supuestos principales y se verificará usando los datos de las encuestas domiciliarias existentes, al igual que un instrumento de encuestas diseñado especialmente para una muestra más reducida y grupos representativos y otras técnicas de PRA. En el análisis se triangulará la información de cada una de estas fuentes de datos.

Análisis de costos-beneficios o eficacia en función de los costos

Aunque este tipo de análisis no se relaciona estrictamente con la medición de los efectos, permite que las autoridades responsables midan la eficiencia de los programas comparando intervenciones alternativas sobre la base del costo de producir un resultado determinado. Puede aumentar enormemente las implicaciones de política de la evaluación de impacto y, por lo tanto, también se debería incluir en el diseño de cualquier evaluación de impacto (un estudio más completo del análisis de costos-beneficios y eficacia en función de los costos se encuentra en *Handbook on Economic Analysis of Investment Operations*, Banco Mundial, 1996.)

Con el análisis de costos-beneficios se intenta medir, en términos monetarios, la eficiencia económica de los costos de un programa versus sus beneficios. Para muchos proyectos, especialmente en los sectores sociales, no es posible medir todos los beneficios en términos monetarios. Por ejemplo, los beneficios de que un programa proporcione insumos escolares (libros de texto, mobiliario para la sala, programas preescolares) serían el aumento del aprendizaje. En lugar de medir los resultados monetarios, se podrían usar las puntuaciones del progreso en materia de aprendizaje para cuantificar los beneficios. Para ello se necesitaría realizar un análisis de la eficacia en función de los costos. Los conceptos para ambos tipos de análisis son los mismos.

Los principales pasos del análisis de costos-beneficios y eficacia en función de los costos son la identificación de todos los costos y beneficios del proyecto y luego calcular una razón costo-eficacia. Al calcular los costos, se debe incluir el valor de la intervención misma, al igual que todos los demás costos, como costos administrativos, de entrega, inversión (actualizados al valor actual neto), el valor monetario de los bienes o servicios proporcionados en forma gratuita, los costos sociales como el deterioro ambiental y los

riesgos para la salud. Los beneficios pueden ser monetarios, como el aumento en el ingreso, o bien el número de unidades entregadas, valoraciones de las pruebas o mejoramientos de salud. Cuando no se puedan cuantificar los beneficios, será posible usar indicadores subjetivos, como sistemas de clasificación o ponderación. Sin embargo, este enfoque puede ser complicado para interpretar valoraciones subjetivas.

Una vez determinados los costos y beneficios, la razón eficacia en función de los costos (R) es entonces $R = \text{costo/unidad (o beneficio)}$. Esta razón se puede comparar en todas las intervenciones para medir la eficiencia. En teoría, esta técnica es bastante directa. En la práctica, sin embargo, hay muchos riesgos involucrados en la identificación y cuantificación de los costos y beneficios. Es importante asegurarse que se seleccionen los indicadores adecuados, que las metodologías y supuestos económicos usados sean coherentes en todas las razones y que las razones de hecho sean comparables. Y al igual que con otras técnicas que se usan en el análisis de los efectos, la mejor forma de realizar la medición de la eficacia en función de los costos es cuando ésta se incluye en el diseño de la evaluación desde las primeras etapas. Esto permite recopilar la información necesaria sobre costos y beneficios y asegurar la coherencia.

Elección de una metodología

Dada la multiplicidad de los tipos de proyectos, preguntas de la evaluación, disponibilidad de datos, restricciones de tiempo y circunstancias del país, cada estudio de evaluación de impacto será diferente y requerirá cierta combinación de las metodologías adecuadas, tanto cuantitativas como cualitativas. El evaluador debe examinar con cuidado las opciones metodológicas al diseñar el estudio, con el objetivo de producir los resultados más sólidos posibles. Entre los métodos cuantitativos, se considera que los diseños experimentales son el enfoque óptimo y las comparaciones pareadas una alternativa subóptima. Sin embargo, otras técnicas también puede producir resultados confiables, especialmente con un buen diseño de evaluación y datos de alta calidad.

La evidencia de las evaluaciones de “práctica óptima” revisadas para este manual pone de relieve que las metodologías elegidas para la evaluación del impacto no son mutuamente excluyentes. Desde luego, en las evaluaciones más robustas con frecuencia se combinan métodos para asegurar su solidez y prever las contingencias en su implementación. Desde una perspectiva metodológica, una combinación muy recomendada es unir un enfoque “con y sin” con un enfoque “antes y después” que use datos básicos y de seguimiento (Subbarao y otros, 1999). Contar con datos básicos permitirá que los evaluadores verifiquen la integridad de los grupos de tratamiento y de comparación, evalúen la focalización y preparen una evaluación sólida de los efectos. Esto aplica incluso para los diseños de control aleatorio. Aunque la aleatorización asegura un grupos de tratamiento y de comparación equivalentes en el momento de ésta, este aspecto no debería influir para que evaluadores piensen que no necesitan datos básicos. Por cierto, los datos básicos podrían ser decisivos para reconstruir por qué se producen ciertos hechos y controlar por estos hechos en la evaluación del impacto.

También se insiste en recomendar la incorporación de análisis de costos-beneficios o

eficacia en función de los costos. Esta metodología puede permitir que las autoridades responsables comparen intervenciones alternativas sobre la base del costo de producir un resultado determinado. Esto es de particular importancia en el marco de los países en desarrollo, en que los recursos son extremadamente limitados.

Por último, es ideal combinar métodos cuantitativos y cualitativos, porque esto proporcionará el efecto cuantificable de un proyecto y una explicación de los procesos e intervenciones que originaron estos resultados. Aunque cada evaluación del impacto tendrá características únicas que requieren diferentes enfoques metodológicos, algunas cualidades generales de una evaluación del impacto de práctica óptima incluyen:

- Una estimación del escenario contrafactual realizada (a) usando una asignación aleatoria para crear un grupo de control (diseño experimental) y (b) usando en forma adecuada y cuidadosa otros métodos como el de pareo para crear un grupo de comparación (diseño cuasi experimental).
- Para controlar las diferencias antes y después del programa en los participantes y establecer los efectos del programa, hay datos pertinentes recopilados como referencia y en el seguimiento (además de un tiempo suficiente que permita observar los efectos del programa).
- Los grupos de tratamiento y de comparación son de tamaño suficiente como para establecer inferencias estadísticas con una mínima eliminación.
- Incluye el análisis de costos-beneficios o eficacia en función de los costos para medir la eficiencia del proyecto.
- Incorpora técnicas cualitativas para permitir la triangulación de los resultados.

Capítulo 2

Pasos clave en el diseño e implementación de una evaluación de impacto *

Emprender un estudio de evaluación del impacto puede ser bastante difícil y costoso, ya que en cada etapa del proceso surgirán problemas para implementarlo. Estas dificultades ponen de relieve la importancia de contar con un estudio bien diseñado, un equipo comprometido y muy calificado y una buena comunicación entre los miembros del equipo de evaluación. Si se incorpora la evaluación en la etapa inicial del diseño de un proyecto, será posible obtener resultados en forma oportuna, de modo que las conclusiones se puedan utilizar para ajustar componentes específicos durante el proyecto.

Sin considerar el tamaño, el tipo de programa ni la metodología usada para la evaluación, se deben seguir varios pasos clave que se resumen a continuación (Recuadro 2.1). En este capítulo se analizarán estos pasos y se examinarán los diversos problemas que pueden surgir en la implementación. La secuencia de estos pasos es decisiva, especialmente para asegurar la recolección de los datos necesarios antes que comience la implementación del proyecto. La planificación oportuna permite aleatorizar, crear comparaciones pareadas ex ante, recolectar datos básicos e identificar encuestas futuras que se podrían usar en un enfoque de correspondencia de puntuación de la propensión.

Durante la identificación y preparación del proyecto, se debe realizar todo el trabajo de diseño y recolección de datos. Es ideal disponer de algunos resultados durante el curso de la implementación del proyecto, de modo que puedan aportar al mejoramiento del diseño del proyecto si es necesario. En el Proyecto de Nutrición y Desarrollo Preescolar de Uganda (véase el Capítulo 4), se ilustra un buen ejemplo de cómo un proyecto incorporó planes de evaluación desde las primeras etapas.

Determinación acerca de si realizar o no una evaluación

Como primer paso, hay que determinar si se requiere o no una evaluación del impacto. Como se señaló más arriba, las evaluaciones del impacto se diferencian de otras evaluaciones en que se centran en examinar la causalidad. Dada la complejidad y el costo de realizar una evaluación del impacto, se deben examinar sus costos y beneficios y considerar si sería más adecuado algún otro método, como el control de indicadores de resultado clave o una evaluación del proceso (enfoques que no se deben considerar como sustitutos de las evaluaciones del impacto, ya que de hecho a menudo aportan componentes decisivos a éstas). Y quizás la información más importante para decidir si llevar a cabo o no una evaluación sea un fuerte apoyo político y financiero.

* Este capítulo se basa en gran medida en un documento preparado por Laura Rawlings, *Implementation Issues in Impact Evaluation*, procesado, 1999.

Recuadro 2.1 Principales pasos en el diseño e implementación de una evaluación de impacto

Durante la identificación y preparación del proyecto

1. Determinar si realizar o no una evaluación
2. Aclarar los objetivos de la evaluación
3. Examinar la disponibilidad de datos
4. Diseñar la evaluación
5. Formar el equipo de evaluación
6. Si se recopilan datos:
 - (a) Diseñar y seleccionar muestras
 - (b) Elaborar instrumentos de recopilación de datos
 - (c) Reunir y capacitar personal para trabajo en terreno
 - (d) Realizar pruebas piloto
 - (e) Recopilar datos
 - (f) Administrar datos y acceder a ellos

Durante la implementación del proyecto

7. Recopilar datos actuales
8. Analizar los datos
9. Redactar los resultados y analizarlos con las autoridades responsables y otras partes interesadas
10. Incorporar los resultados en el diseño del proyecto

Es más factible movilizar el esfuerzo y los recursos adicionales necesarios para realizar la evaluación del impacto cuando el proyecto es innovador, se puede repetir, supone importantes asignaciones de recursos y comprende intervenciones bien definidas. Por ejemplo, la evaluación de impacto del Fondo de Inversión Social boliviano cumplía con todos estos criterios. En primer lugar, se consideró que el nuevo modelo de fondo social introducido en Bolivia era innovador y susceptible de repetir; en segundo lugar, el fondo social ha tenido a cargo alrededor de un 25% de todas las inversiones públicas del país desde el inicio de la evaluación y, en tercer lugar, las intervenciones estaban bien definidas por el menú de subproyectos del fondo social.

También se deben priorizar la evaluación del impacto si en el marco del proyecto en cuestión se está lanzando un nuevo enfoque, como un programa piloto que luego se considerará extender basándose en los resultados de la evaluación, o los Préstamos para el Aprendizaje e Innovación del Banco Mundial. Este fundamento permitió que la reforma de autonomía escolar de Nicaragua fuera una buena candidata para una evaluación de impacto. El estudio de evaluación siguió las pruebas del gobierno de un nuevo modelo de

administración escolar descentralizada desde su etapa piloto a mediados de los años noventa hasta su expansión hacia casi todas las escuelas secundarias y, en la actualidad, aproximadamente la mitad de todas las escuelas primarias. El equipo internacional rigurosamente coordinado que manejó la evaluación incluía personal local de la unidad de investigación y evaluación del Ministerio de Educación y la oficina de coordinación de Proyectos de Educación Primaria del Banco Mundial en Managua. Su participación garantizaba que la evaluación daría cuenta de decisiones clave de políticas con respecto a la modificación y expansión del programa piloto.

Otra importante consideración es asegurar que el programa que se evaluará se encuentre en una etapa suficientemente avanzada para ser sometido a una evaluación de impacto. Los proyectos piloto y las reformas incipientes con frecuencia tienden a sufrir modificaciones de su contenido y en cuanto a cómo, cuándo y por quién serán implementados. Estos cambios pueden perjudicar la coherencia del esfuerzo de evaluación, especialmente los diseños experimentales y otros tipos de evaluaciones prospectivas que se basan en datos básicos y de seguimiento de grupos de tratamiento y de control claramente establecidos. Cuando las políticas que se evaluarán aún se están definiendo, podría ser recomendable no usar una evaluación del impacto, para así otorgar flexibilidad al proyecto.

Obtener el apoyo de las autoridades responsables y de los financistas para una evaluación del impacto puede ser difícil, pero es un requisito previo para seguir adelante. Deben estar convencidos de que la evaluación es un ejercicio útil para abordar preguntas que serán pertinentes a las decisiones relacionadas con el perfeccionamiento, expansión o reducción del programa evaluado. También deben estar convencidos de la legitimidad del diseño de la evaluación, y por consiguiente, de los resultados, en particular cuando éstos no son tan positivos como se esperaba.

Financiar una evaluación del impacto sigue siendo un problema difícil para los administradores de programas, al igual que para las contrapartidas de los clientes. El problema financiero se ve agravado por la dificultad general de obtener los datos sobre costos de la evaluación. Y quizás el problema más engorroso surge de la disposición de fondos públicos para la evaluación: si los resultados de la evaluación se usarán para guiar políticas que se aplicarán fuera de las fronteras nacionales dentro de las cuales se realiza la evaluación, como a menudo es el caso, ¿por qué un país habría de asumir los costos de la evaluación? Entre los estudios de casos que contaban con información sobre fuentes de financiamiento, ésta indica que los países con frecuencia asumen la mayor parte de los costos de la evaluación, pero no su totalidad. Como se analiza con más detalle en el Capítulo 4, muchos de los casos revisados sugieren que para implementar con éxito una evaluación del impacto no sólo se requiere un compromiso importante por parte de los países solicitantes, sino también la participación de personal del Banco Mundial o de investigadores y asesores externos, necesitando recursos adicionales a los que proporciona el país.

Aclaración de los objetivos de la evaluación

Una vez que se ha determinado que la evaluación del impacto es adecuada y se justifica,

establecer objetivos claros y acordar los puntos básicos en que se centrará la evaluación de antemano contribuirá enormemente a su éxito. Los objetivos claros son esenciales para identificar las necesidades de información, definir los indicadores de resultados y efectos y crear una estrategia de evaluación sólida que proporcione respuestas a las preguntas planteadas. La utilización de un planteamiento de marco lógico (log) proporciona una herramienta adecuada y comúnmente usada para identificar las metas del proyecto y las necesidades de información sobre cuyas bases se creará la evaluación.

El marco lógico, cada vez usado en el Banco Mundial, se basa en una simple matriz cuatro por cuatro en que se asocia información sobre los objetivos de un proyecto con la forma en que se realizará el seguimiento de los resultados usando puntos de referencia y planes de trabajo, el efecto que tendrán los resultados del proyecto en una institución o sistema beneficiario, a la forma en que éste se medirá y a la forma en que se usarán los aportes de información para entregar resultados (ejemplos de esto se encuentran en el Anexo 5). En otras palabras, se supone que el efecto previsto del proyecto es una función de los resultados del proyecto, al igual que de una serie de otros factores. A su vez, los resultados son una función de la información del proyecto y de otros factores ajenos al proyecto. Luego se deben identificar medidas cuantificables para cada vínculo en el ciclo del proyecto. Este enfoque no impide que el evaluador también examine los efectos no previstos de un proyecto, pero sirve para mantener claros y focalizados los objetivos de la evaluación. Las técnicas cualitativas también son útiles para obtener participación en la aclaración de los objetivos de la evaluación y los indicadores de impacto resultantes.

Aunque aparentemente el planteamiento del objetivo pareciera ser una de las partes más fáciles del proceso de evaluación, puede ser en extremo difícil. Por ejemplo, los planteamientos que son demasiado amplios no se prestan para ser sometidos a una evaluación. El planteamiento del objetivo de la evaluación del PROBECAT de México (Anexo 1.9), de que la evaluación busca examinar el “efecto del programa de capacitación PROBECAT en los resultados del mercado laboral”, sería más preciso si se limitara al efecto de PROBECAT en las horas trabajadas, los ingresos por hora, el salario mensual y el tiempo para conseguir el primer empleo para diferentes tipos de trabajadores. La evaluación de PROGESA de México es un buen ejemplo de la creación de un esquema claro y el delineamiento de múltiples objetivos desde el comienzo, con un análisis separado para cada componente y con los objetivos detallados en subcategorías (Anexo 1.10). Esto fue especialmente importante, porque la intervención fue bastante compleja al tener que abordar no sólo el efecto del programa, sino también aspectos de la focalización y secuencia cronológica de las operaciones del programa.

Revisar otros componentes de evaluación, como la eficacia en función de los costos o las evaluaciones del proceso, también pueden ser importantes objetivos de un estudio y pueden complementar la evaluación del impacto. La eficacia en función de los costos puede constituir una preocupación muy especial para las autoridades responsables, cuyas decisiones serán reducir, expandir o reformar la intervención que se está evaluando. En temas relacionados con la entrega de servicios, una evaluación de los procesos podría ser pertinente para evaluar los procedimientos, dinámica, normas y restricciones bajo los cuales se lleva a cabo un determinado programa.

Examen de la disponibilidad de datos

Se pueden emplear muchos tipos de datos para realizar estudios de evaluación del impacto. Éstos pueden incluir una gama que fluctúa desde encuestas transversales o de panel hasta entrevistas cualitativas flexibles. Es ideal que esta información exista a nivel individual, para asegurar que se pueda evaluar el verdadero efecto. La información a nivel de hogares puede ocultar una asignación intrafamiliar de los recursos que afecta a las mujeres y niños, porque con frecuencia tienen un acceso más limitado a los recursos productivos del hogar. En muchos casos, la evaluación del impacto aprovechará cierto tipo de datos existentes o se apoyará en alguna encuesta en curso, lo que puede significar un considerable ahorro de costos. Sin embargo, con este enfoque pueden surgir problemas de secuencia cronológica del esfuerzo de recolección de datos y con la flexibilidad del diseño del cuestionario. En el Recuadro 2.2 se destacan algunos puntos clave para recordar al examinar el uso de los recursos de datos existentes para la evaluación del impacto.

Con cierta creatividad, podría ser posible maximizar los recursos de información existentes. Un buen ejemplo es la evaluación del Fondo de Inversión Social Hondureño (véase el Capítulo 4). En ese estudio se utilizó un módulo de la encuesta nacional de ingresos y gastos en el cuestionario de fondos sociales, lo que permitió comparar el ingreso de los beneficiarios del fondo social con las medidas nacionales para evaluar la focalización de la pobreza (Walker y otros, 1999).

Al nivel más básico, se requerirán datos sobre el universo de la población de interés como base para determinar los tamaños de las muestras, crear el marco de muestreo y seleccionar la muestra. Otros tipos de datos que pueden estar disponibles en un país determinado y que se pueden emplear para distintas evaluaciones de impacto incluyen (ver Valadez y Bamberger, 1994) encuestas ingresos y gastos familiares, Encuestas Sobre las Condiciones de Vida (ENCOVI); encuestas del mercado laboral, registros de cooperativas, asociaciones de crédito y otras instituciones financieras, registros escolares sobre asistencia, repetición y rendimiento en exámenes; registros de salud pública sobre mortalidad infantil, incidencia de diferentes enfermedades infecciosas, número de mujeres que busca orientación sobre anticoncepción o consumo de preservativos, encuestas especializadas aplicadas por universidades, organizaciones no gubernamentales (ONG) y grupos asesores, vigilancia de datos de administradores de programas y estudios de casos de proyectos.

Recuadro 2.2 Puntos clave para identificar las fuentes de datos disponibles para la evaluación de impacto

- Conocer bien el programa. Es arriesgado iniciar una evaluación sin conocer bastante sobre los detalles administrativos e institucionales del programa; esta información en general proviene de la administración del programa.
- Recopilar información sobre los “hechos estilizados” pertinentes al entorno. Los hechos pertinentes podrían incluir el mapa de pobreza, la forma en que funciona el mercado laboral, las principales divisiones étnicas y otros programa públicos pertinentes.
- Ser ecléctico con respecto a los datos. Las fuentes pueden abarcar entrevistas informales no estructuradas con participantes en el programa y datos cuantitativos de muestras representativas. Sin embargo, es muy difícil plantear preguntas de un escenario contrafactual en las entrevistas o grupos representativos; se debe intentar preguntar a alguien que actualmente participe en un programa público: “¿Qué estaría haciendo ahora si este programa no existiera?” Puede ser valioso conversar con los participantes del programa, pero es poco probable que ello por sí sólo proporcione una evaluación creíble.
- Asegurarse de que existen datos sobre los indicadores de resultados y variables explicativas pertinentes. Estas últimas deben abordar la heterogeneidad en los resultados basada en la participación en el programa. Dependiendo, por ejemplo, si se cuenta con cierto nivel de instrucción, los resultados pueden ser diferentes. Podría ser imposible observar el efecto del programa a menos que se controle por esa heterogeneidad.
- Dependiendo de los métodos usados, también se podrían necesitar datos sobre variables que influyen en la participación, pero no influyen en los resultados dada la participación. Estas variables instrumentales pueden ser valiosas para separar los efectos probablemente causales de los programas no aleatorios (Recuadro 1.2).
- Los datos sobre resultados y otras variables explicativas pertinentes pueden ser cuantitativos o cualitativos. Pero debe ser posible organizar la información en cierto tipo de estructura sistemática de datos. Un ejemplo simple y común es que se tienen valores de diversas variables, incluido uno o más indicadores de resultados para diversas unidades de observación (individuos, hogares, empresas, comunidades).
- Frecuentemente las variables sobre las que se tienen datos y las unidades de observación que se usan se seleccionan como parte del método de evaluación. Estas elecciones deben estar vinculadas al conocimiento previo sobre el programa (desde luego, sus objetivos, pero también cómo se maneja) y el escenario en que se introdujo.
- La fuente específica de los datos sobre los resultados y sus determinantes, incluida la participación en el programa, en general proviene de datos de encuestas de algún tipo. La unidad de observación podría ser el hogar, empresa o área geográfica, dependiendo del tipo de programa que se está estudiando.
- Los datos de las encuestas con frecuencia se pueden complementar con otros datos útiles sobre el programa (como de la base de datos de supervisión del proyecto) o escenario (como de las bases de datos geográficas).

Uso de datos de encuestas existentes. Muchas encuestas también pueden estar en las etapas de planificación o en curso. Si se planifica una encuesta que mida los indicadores requeridos, la evaluación podría tomar una muestra de la población de interés durante el curso de la encuesta general (por ejemplo, para usarla para el método de correspondencia de puntuación de la propensión), como se hizo con la evaluación del Fondo de Inversión Social de Nicaragua y la evaluación del programa de empleo con salario mínimo TRABAJAR de Argentina (Jalan y Ravallion, 1998). A su vez, si se planifica una encuesta que abarcará la población de interés, la evaluación debe poder introducir una pregunta o serie de preguntas como parte de la encuesta o agregar una encuesta cualitativa para complementar la información cuantitativa. Por ejemplo, el programa Crédito con Educación de Ghana incluyó un conjunto de entrevistas cualitativas a partes interesadas clave y a grupos representativos participantes y no participantes que confirmaron cualitativamente los resultados cuantitativos (Anexo 1.6). La evaluación examinó el efecto del programa en el estado nutricional y la seguridad alimentaria de los hogares pobres. Los datos cuantitativos incluyeron preguntas específicas sobre ingreso y gasto familiar y nivel de conocimientos, mientras que los datos cualitativos se centraron en el estado de habilitación y toma de decisiones de las mujeres en el hogar, las redes sociales, la confianza en sí mismas, etc.

Diseño de la evaluación

Una vez que los objetivos y datos disponibles estén claros, es posible comenzar con la fase de diseño del estudio de la evaluación del impacto. La elección de metodologías dependerá de las preguntas, secuencia cronológica, restricciones presupuestarias y capacidad de implementación de la evaluación. Se deben contrapesar los pros y contras de los distintos tipos de diseño analizados en el Capítulo 1, para determinar las metodologías que son más adecuadas y las técnicas cuantitativas y cualitativas que se pueden integrar para complementarse entre sí.

Incluso después que se ha determinado el diseño de la evaluación y se ha incorporado al proyecto, los evaluadores deben estar preparados para ser flexibles y modificar el diseño a medida que se implementa el proyecto. Además, debe disponerse el seguimiento de las intervenciones del proyecto si la evaluación incluye datos básicos y de seguimiento, de modo que el esfuerzo de evaluación coincida con la marcha real del proyecto.

Al definir el diseño, también es importante determinar cómo la evaluación del impacto se ajustará a la estrategia de control y evaluación más amplia aplicada a un proyecto. Se deben supervisar todos los proyectos, de modo que los administradores, prestamistas y autoridades responsables puedan seguir el desarrollo del proyecto. Como se señaló anteriormente, el esfuerzo de evaluación debe ajustarse a las necesidades de información del proyecto.

Preguntas de evaluación. Las preguntas de evaluación que se plantean están bastante ligadas al diseño de la evaluación en términos del tipo de datos recopilados, la unidad de análisis, las metodologías usadas y la secuencia cronológica de las diversas etapas. Por

ejemplo, al evaluar el impacto de los libros de textos en los resultados de aprendizaje, sería necesario adaptar la evaluación para medir el efecto en los estudiantes, salas de clases y maestros durante un año escolar determinado. Esto sería muy distinto a medir el efecto de los servicios entregados a través de inversiones en fondos sociales, en que se requerirían datos sobre los hogares y la infraestructura social. Los estudios de casos del Anexo I proporcionan los demás ejemplos sobre cómo la interrogante de la evaluación puede afectar el diseño de la misma.

Al aclarar las preguntas de la evaluación, también es importante considerar las implicancias de género del efecto del proyecto. Al principio esto no siempre es evidente; sin embargo, en la implementación del proyecto podrían haber efectos secundarios en el hogar que no necesariamente se captan sin una recolección de datos ni esfuerzos de análisis específicos.

Asuntos de secuencia cronológica y presupuesto. El tema más crucial de la secuencia cronológica se relaciona con la posibilidad de comenzar el diseño de la evaluación antes que se implemente el proyecto y el momento en que se necesiten los resultados. También es útil identificar de antemano en qué puntos durante el ciclo del proyecto se necesitará información del esfuerzo de evaluación, de modo que se pueda vincular la recolección de datos y las actividades de análisis. Tener los resultados en forma oportuna puede ser decisivo para las decisiones de políticas —por ejemplo, durante una revisión del proyecto, cerca de un período de elecciones o cuando se tomen decisiones con respecto a la continuación del proyecto.

Algunos métodos demoran más en implementarse que otros. La asignación aleatoria y los métodos antes y después (por ejemplo, comparaciones reflexivas) tardan más que los enfoques de comparación pareada *ex post*. Al usar enfoques antes y después que utilizan evaluaciones básicas (de referencia) y seguimiento, se debe disponer de suficiente tiempo para que el último miembro del grupo de tratamiento reciba la intervención y luego por lo general de más tiempo para que se materialicen y se observen los efectos posteriores al programa. Grossman (1994) sugiere que en general hay que dejar transcurrir entre 12 y 18 meses después de incorporar la muestra en la intervención para examinar los efectos. En los proyectos del Banco Mundial con niveles de referencia, la espera para que se implemente la intervención y que se materialicen los resultados puede tomar años. En la evaluación del Fondo de Inversión Social de Bolivia, por ejemplo, que se basó en datos básicos recopilados en 1993, los datos de seguimiento recién se recolectaron en 1998 debido al tiempo necesario para llevar a cabo las intervenciones (proyectos sanitarios y de suministro de agua, consultorios médicos y escuelas) y se materializaran los efectos a nivel de salud y educación en la población beneficiaria. Un período similar se requirió para evaluar un proyecto de educación primaria en Pakistán en que se utilizó un diseño experimental con encuestas básicas y de seguimiento para evaluar el impacto de las escuelas de las comunidades en los resultados de los alumnos, incluido el rendimiento académico.

Las exigencias de la secuencia cronológica de la evaluación no pueden constituirse en el motor del proyecto que se está evaluando. Por su misma naturaleza, las evaluaciones están sujetas al marco cronológico establecido por el resto del proyecto. Las evaluaciones deben esperar en el caso de proyectos de lento desembolso y generación de intervenciones. Y aun

cuando los proyectos avancen al ritmo establecido, algunas intervenciones tardan más en llevarse a cabo, como los proyectos de infraestructura. El marco cronológico de la evaluación también es sensible a los indicadores seleccionados, porque muchos, como las variaciones en las tasas de fecundidad o el rendimiento educacional, tardan más en manifestarse en la población beneficiaria.

Capacidad de implementación. Una consideración final en la escala y complejidad del diseño de la evaluación es la capacidad de implementación del equipo de evaluación. Los aspectos de la implementación pueden ser bastante difíciles, especialmente en los países en desarrollo, donde hay poca experiencia en las evaluaciones de programas e investigaciones aplicadas. La composición del equipo de evaluación es muy importante, al igual que la experiencia de los miembros de equipo en diferentes tipos de metodologías y su capacidad relativa a otras actividades que realice la unidad de evaluación. Esto se aplica en particular cuando se trabaja con organismos del sector público con múltiples responsabilidades y limitado personal. Es importante conocer la carga de trabajo de la unidad, para evaluar no sólo la forma en que esto afectará la calidad de la evaluación que se está realizando, sino también el costo de oportunidad de la evaluación con respecto a otros esfuerzos de los cuales está encargada la unidad. Existen varios ejemplos de esfuerzos de evaluación que se vieron frustrados cuando el personal clave fue requerido para otros proyectos y por lo tanto no le fue posible realizar la recolección de datos programada en el momento decisivo (como un momento específico durante el año escolar o durante la temporada agrícola). Estas situaciones se pueden evitar coordinándose con los administradores de la unidad a cargo de la evaluación, para asegurar que se logre un equilibrio con respecto a la secuencia cronológica de las diversas actividades, al igual que la distribución de personal y recursos en estas actividades. Alternativamente, puede ser preferible contratar una empresa privada para que realice la evaluación (como se describe más adelante).

Formación del equipo de evaluación

Se necesitan diversos conocimientos en el trabajo de evaluación. La calidad y utilidad final de la evaluación del impacto se puede mejorar en gran medida si desde un principio los miembros del equipo se coordinan con las autoridades responsables. Por lo tanto, es importante identificar a los miembros del equipo lo antes posible, acordar sus funciones y responsabilidades y establecer mecanismos de comunicación durante los momentos clave de la evaluación.

El equipo básico se compone del administrador o director de la evaluación, analistas (economistas y otros científicos sociales) y, para los diseños de la evaluación que implican la recolección de nuevos datos, un experto en muestreo, un diseñador de encuestas, un administrador del trabajo en terreno y el respectivo equipo, y los administradores y procesadores de datos (una guía global para diseñar e implementar encuestas se encuentra en Grosh y Muñoz, 1996). Según el tamaño, alcance y diseño del estudio, algunas de estas responsabilidades serán compartidas o bien se agregarán otras necesidades de personal a este equipo básico. En los casos en que los analistas de políticas no hayan tenido experiencia en la integración de enfoques cuantitativos y cualitativos, podría ser necesario emplear tiempo adicional en la etapa inicial de la formación del equipo, para sensibilizar a

los miembros del equipo y asegurar una completa colaboración. Las responsabilidades generales de los miembros del equipo incluyen lo siguiente:

- **Administrador de la evaluación** — Es el encargado de establecer las necesidades de información e indicadores para la evaluación (que con frecuencia se determinan con el cliente usando un enfoque de marco lógico), delinear sus términos de referencia, seleccionar la metodología pertinente e identificar el equipo a cargo de la evaluación. En muchos casos, el administrador de la evaluación también llevará a cabo el análisis de políticas.
- **Analistas de política** — Se necesita un economista para el análisis cuantitativo, al igual que un sociólogo o antropólogo para garantizar el aporte participativo y el análisis cualitativo en las diferentes etapas de la evaluación del impacto. Ambos deben participar en la redacción del informe de evaluación.
- **Experto en muestreo** — Puede dirigir el proceso de selección de muestras. Para datos cuantitativos, el experto en muestreo debe poder realizar cálculos de poder para determinar los tamaños adecuados de la muestra para los indicadores establecidos, seleccionar la muestra, revisar los resultados de la muestra real versus la muestra diseñada e incorporar las ponderaciones del muestreo para el análisis. Para los datos cualitativos, el experto en muestreo debe guiar el proceso de selección de la muestra en coordinación con los analistas, asegurando que los procesos establecidos garanticen la selección de los informantes correctos. El experto en muestreo también debe estar a cargo de seleccionar los lugares y grupos para la prueba piloto y con frecuencia necesitará estar en contacto con un coordinador de información local responsable de recolectar datos para el experto en muestra de los cuales se tomará la muestra.
- **Diseñador de encuesta** — Podría ser una persona o un equipo, a cargo de diseñar los instrumentos de recolección de datos, los manuales y libros de códigos adjuntos y coordinarse con el o los administradores de la evaluación para asegurar que los instrumentos de recolección de datos realmente produzcan los datos requeridos para el análisis. Esta persona o equipo debe participar en la prueba piloto y perfeccionar los cuestionarios.
- **Administrador y personal de trabajo en terreno** — El administrador debe ser responsable de supervisar todo el esfuerzo de recolección de datos, desde la planificación de las rutas de recolección de datos hasta formar y programar los equipos de trabajo en terreno, los que en general estarán compuestos de supervisores y entrevistadores. Los supervisores generalmente administran el personal del trabajo en terreno (habitualmente entrevistadores, operadores de ingreso de datos y conductores) y son responsables de la calidad de los datos recolectados en terreno. Los entrevistadores aplican los cuestionarios. En algunas culturas, es necesario asegurar que las encuestas sean realizadas por entrevistadores de ambos sexos y que se apliquen en forma separada a hombres y mujeres.
- **Administradores y procesadores de datos** — Estos miembros del equipo diseñan los

programas de ingreso de datos, ingresan los datos, comprueban su validez, proporcionan la documentación necesaria de los datos y producen resultados básicos que pueden ser verificados por los analistas de datos.

Al formar el equipo de evaluación, también hay algunas decisiones importantes que el administrador de la evaluación debe tomar acerca de la capacidad local y las disposiciones institucionales adecuadas para garantizar la imparcialidad y la calidad de los resultados de la evaluación. La primera de ellas es si existe capacidad local para implementar la evaluación o parte de ella y qué tipo de supervisión y asistencia exterior se necesitará. La capacidad de evaluación varía enormemente de un país a otro. Y aunque son cada vez más comunes los contratos internacionales que permiten que las empresas de un país realicen evaluaciones en otro país (un ejemplo es la evaluación Progresá realizada por el International Food and Policy Research Institute), la práctica general de los proyectos apoyados por el Banco Mundial parece ser que la evaluación se implemente usando personal local, con el debido apoyo de una considerable supervisión internacional. Por lo tanto, es necesario evaluar en forma crítica la capacidad local y determinar quién estará a cargo de qué aspectos del esfuerzo de evaluación. Sin considerar la composición final del equipo, es importante designar un administrador de evaluación que pueda trabajar en forma eficaz con los productores de datos y con los analistas y autoridades responsables que utilizarán los resultados de la evaluación. Si esta persona no está instalada en el país, es recomendable que se designe a un administrador local para que coordine el esfuerzo de evaluación junto con el administrador internacional.

La segunda decisión es si trabajar con una empresa privada o con un organismo público. Las empresas privadas pueden ser más confiables con respecto a la entrega de resultados en forma oportuna, pero se pierde la creación de capacidades en el sector público. Además, las empresas privadas, comprensiblemente, a menudo están menos dispuestas a incorporar elementos a la evaluación que harán más costoso el esfuerzo. Cualquiera sea la contrapartida o combinación de contrapartidas que finalmente se establezca, es esencial hacer una revisión sólida de las actividades de evaluación anteriores de los posibles colaboradores para tomar una decisión fundamentada.

Y la tercera es el grado de separación institucional que habrá que establecer entre los proveedores de la evaluación y los usuarios de ésta. Hay muchas ventajas en la objetividad proporcionada al realizar la evaluación en forma independiente de la institución responsable del proyecto que se está evaluando. Sin embargo, las evaluaciones a menudo tienen múltiples objetivos, como crear capacidad de evaluación dentro de los organismos gubernamentales y sensibilizar a los operadores de un programa con las realidades de sus proyectos una vez que estos se aplican en terreno. Como mínimo, los usuarios de una evaluación, que pueden abarcar desde autoridades responsables de los organismos gubernamentales de los países solicitantes hasta organizaciones ONG, donantes bilaterales e instituciones de desarrollo internacional, deben participar en la evaluación lo suficiente como para garantizar que se reconozca la legitimidad del proceso de evaluación y que los resultados producidos sean pertinentes para sus necesidades de información. De lo contrario, será menos probable que se usen los resultados de la evaluación para guiar las políticas. En el análisis final, el administrador de evaluación y su o sus clientes deben lograr

el equilibrio correcto entre involucrar a los usuarios de las evaluaciones y mantener la objetividad y legitimidad de los resultados.

Elaboración de datos

Tener datos adecuados y confiables es un aporte básico necesario para evaluar el impacto de un proyecto. Contar con datos de alta calidad es esencial para la validez de los resultados de la evaluación. Como se analizó anteriormente, evaluar los datos que existen es un primer paso importante antes de iniciar cualquier nuevo esfuerzo de recopilación de datos. En el Cuadro 2.1 se relacionan las metodologías de evaluación básicas con las necesidades de datos. En la mayoría de estas metodologías se pueden incorporar técnicas cualitativas y participativas en el diseño del instrumento de encuestas, la identificación de los indicadores y los aportes a la identificación de controles, las variables usadas para la correspondencia o las variables instrumentales.

Cuadro 2.1 Métodos de evaluación y las correspondientes necesidades de datos

<i>Método</i>	<i>Necesidad de datos</i>		<i>Uso de enfoque cualitativo</i>
	<i>Mínima</i>	<i>Ideal</i>	
Controles experimentales o aleatorizados	Corte transversal de un único proyecto con y sin beneficiarios	Encuestas básicas y de seguimiento de los beneficiarios y no beneficiarios. Permite controlar sucesos contemporáneos, además de proporcionar control para medir el efecto. (Esto permite una estimación de diferencia en las diferencias).	<ul style="list-style-type: none"> • Guiar el diseño del instrumento de la encuesta, muestreo • Identificar los indicadores • Recopilar y registrar datos con la ayuda de <ul style="list-style-type: none"> – Datos textuales – Entrevistas informales o semi estructuradas – Reuniones de los grupos representativos o de la comunidad – Observación directa – Métodos participativos – Fotografías – Triangulación – Análisis de datos
Diseños no experimentales	Encuesta extensa, censo, presupuesto nacional o tipo ENCOVI de encuesta en que se muestrea a los beneficiarios	Encuesta extensa y encuesta domiciliaria más reducida basada en el proyecto, ambas con dos momento en el tiempo para controlar por los sucesos contemporáneos	
a) Controles o pareo creados			
b) Comparaciones reflexivas y doble diferencia	Básica y de seguimiento de los beneficiarios	Series cronológicas o panel de los beneficiarios y no	

c) Control estadístico o variable instrumental	Datos transversales representativos de la población beneficiaria con las correspondientes variables instrumentales	beneficiarios comparables Corte transversal y series cronológicas representativas de la población beneficiaria y no beneficiaria, con las variables instrumentales correspondientes
--	--	--

Fuentes: Adaptado de Ezemenari, Rudqvist y Subbarao (1999) y Bamberger

En el caso de evaluaciones que generen sus propios datos, se requerirán los pasos decisivos de diseñar los instrumentos de recolección de datos, muestreo, trabajo en terreno, manejo de datos y acceso a datos. En esta sección no se indica el proceso paso a paso de la forma de realizar una encuesta, sino más bien se proporciona un breve análisis de estos pasos. Parte de la descripción de esta sección, en especial con respecto al muestreo y el manejo de datos, es más pertinente para las evaluaciones basadas en la recopilación y análisis de encuestas de muestras a mayor escala usando datos cuantitativos, que para las evaluaciones que se basan en datos cualitativos y muestras pequeñas.

Decisión sobre lo que se medirá. Al planificar la evaluación, se deben establecer los principales indicadores de resultados y efectos, en lo posible como parte de un enfoque de marco lógico. Para garantizar que la evaluación puede examinar los resultados durante un período de tiempo que responde a las necesidades de las autoridades responsables, se puede definir una jerarquía de indicadores que fluctúe desde indicadores de impacto a corto plazo, como asistencia escolar, hasta indicadores a largo plazo, como rendimiento estudiantil. Con esto se garantiza que aun cuando no se captan inicialmente los efectos finales, se podrán evaluar los resultados del programa. Además, el evaluador debe pensar en medir la entrega de la intervención y tener en cuenta los factores exógenos que puedan tener algún efecto en el resultado de interés.

Los administradores de la evaluación también pueden planificar la aplicación de ésta en diversos períodos, lo que permitirá captar antes los efectos más inmediatos y seguir evaluando las medidas de los resultados finales. Esto se realizó en la evaluación de la Reforma Escolar de Nicaragua, en la cual se estableció el efecto a más corto plazo de la reforma en la participación de los padres y en la asistencia de alumnos y maestros y aún están en evaluación los efectos a más largo plazo en el rendimiento de los alumnos.

También se podría considerar alguna información sobre características de la población beneficiaria no relacionadas estrictamente con la evaluación del impacto, pero de interés para el análisis, como su nivel de pobreza o su opinión acerca del programa. Además, el evaluador también podría desear incluir medidas de costos para realizar algún análisis de eficacia en función de los costos u otras evaluaciones complementarias que no diga relación estricta con la evaluación del impacto.

El tipo de diseño seleccionado para la evaluación del impacto también influirá en las necesidades de datos. Éstas serán específicas para la metodología, población de interés, medidas de los efectos y otros elementos de la evaluación. Por ejemplo, si se usa un enfoque de variable instrumental (uno de los tipos de estrategias de comparación pareada), se deberán identificar e incluir en la recolección de datos la o las variables que servirán de instrumento para separar la participación en el programa de las medidas de resultados. Esto es lo que se hizo en la evaluación de impacto del Fondo de Inversión Social de Bolivia, donde se usó el conocimiento del fondo social y la presencia de ONG como variables instrumentales para evaluar el impacto de las intervenciones del fondo social.

Puede ser útil elaborar una matriz para la evaluación en que se indique la pregunta de interés, los indicadores de resultado que se usarán para evaluar los resultados, la variable y la fuente de datos para la variable. Luego, esta matriz se puede usar para revisar cuestionarios y planificar el trabajo analítico, como se hizo en la evaluación del Fondo de Inversión Social de Emergencia de Nicaragua (véase el Anexo 6).

Elaboración de instrumentos y métodos de recopilación de datos. Elaborar los instrumentos de recolección de datos apropiados que generen los datos necesarios para responder las preguntas de la evaluación puede ser engorroso. Para ello, los analistas deberán participar en la elaboración de las preguntas, la prueba piloto y la revisión de los datos de la prueba piloto. Puede ser decisivo para la calidad de información recopilada que durante la elaboración de los instrumentos participen el administrador en terreno y el administrador de datos, además del personal local, de preferencia analistas que puedan aportar con conocimientos del país y del programa (Grosh y Muñoz, 1996). También es importante asegurar que los datos reunidos se puedan desglosar por género, para así examinar el efecto diferencial de políticas y programas específicos.

En las evaluaciones cuantitativas por lo general se recopila y registra la información en forma numérica o como categorías precodificadas. En las evaluaciones cualitativas, la información en general se presenta como un texto descriptivo con una categorización reducida o ninguna. La información podría incluir las respuestas de los individuos a preguntas de la entrevista que admiten más de una respuesta, notas tomadas en reuniones de grupos representativos o las observaciones de los hechos por parte del evaluador. En algunos estudios cualitativos también se utiliza la clasificación precodificada de datos (Bamberger, 2000). En el Cuadro 2.2 se resume el abanico de instrumentos de recolección de datos y sus fortalezas y debilidades, de los cuales la técnica más comúnmente usada son los cuestionarios.

Las respuestas de los cuestionarios de encuestas pueden ser muy sensibles al diseño; por lo tanto, es importante asegurarse de que la estructura y el formato sean adecuados y que de preferencia los realice personal con experiencia. Por ejemplo, la utilidad de los datos cuantitativos con frecuencia se ha visto severamente obstaculizada por motivos simplemente mecánicos, como la imposibilidad de vincular los datos de una fuente a otra. Esto sucedió en una evaluación de educación nacional en un país en que no se pudieron vincular los datos de los antecedentes de los estudiantes con los resultados de las

calificaciones de las pruebas, lo que hizo imposible evaluar la influencia de las características de los estudiantes en el rendimiento y clasificar las calificaciones de las pruebas según edad, género, nivel socioeconómico o antecedentes educacionales de los alumnos.

Para la recolección de datos cualitativos y cuantitativos, se debe capacitar incluso a personal con experiencia para recopilar los datos específicos para la evaluación. Además, cualquier recolección de datos debe estar guiada por un conjunto de manuales que se puede usar como orientación durante la capacitación y como referencia durante el trabajo en terreno. Según la complejidad de la tarea de recolección de datos, los ejemplos de casos demuestran que la capacitación puede demorar entre tres días y varias semanas.

Realizar pruebas piloto es un paso esencial, porque revelará si el instrumento puede producir de manera confiable los datos requeridos y la forma en que se pueden poner en funcionamiento los procedimientos de recolección de datos. Una prueba piloto debe asemejarse al trabajo en terreno real lo más exactamente posible. Por este motivo, es útil tener programas de ingreso de datos preparados al momento de realizar la prueba, para verificar su funcionalidad e incluir en el trabajo en terreno real los resultados de la prueba en las distintas poblaciones y áreas geográficas.

Muestreo. El muestreo es un arte que de preferencia lo debe practicar un especialista experimentado en muestreo. El diseño no necesita ser complicado, pero los conocimientos prácticos del especialista en muestreo deben guiar a éste en la determinación de los marcos de muestreo y estrategias de selección adecuados (este análisis sobre muestreo se refiere principalmente a temas relacionados con evaluaciones en que se recopilan datos cuantitativos de muestras estadísticamente representativas de mayor tamaño). El especialista en muestreo debe ser incorporado al proceso de evaluación desde las primeras etapas, para revisar la información disponible necesaria para seleccionar la muestra y determinar si se requerirá algún trabajo de enumeración, que puede requerir mucho tiempo.

Como con otras partes del trabajo de evaluación, es importante que exista coordinación entre el especialista en muestreo y el equipo de evaluación. Esto es especialmente decisivo al realizar comparaciones pareadas, porque aquí el diseño del muestreo es la base del "pareo", que es la esencia del diseño de evaluación y la creación del escenario contrafactual. En estos casos, el especialista en muestreo debe trabajar de cerca con el equipo de evaluación al elaborar los criterios que se aplicarán, para que correspondan a los grupos de tratamiento y de comparación. En la evaluación del proyecto de reforma de autonomía escolar de Nicaragua, por ejemplo, se estratificaron las escuelas autónomas por tipo de escuela, matrícula, tiempo en la reforma y ubicación y se asociaron a una muestra de escuelas no autónomas usando las mismas estratificaciones, excepto el tiempo en la reforma. Esto se puede facilitar si se asigna un miembro del equipo encargado del trabajo de recolección de datos para que asista al especialista en la obtención de la información requerida, incluidos los datos sobre los indicadores de resultado seleccionados para los cálculos de poder (una estimación del tamaño de la muestra requerido para probar la significancia estadística entre dos grupos), una lista de la población de interés para la selección de la muestra y detalles sobre las características de los posibles grupos de

tratamiento y de comparación que sean importantes para el proceso de selección de la muestra.

Existen muchas relaciones inversas de compensación entre costos y exactitud del muestreo que deben aclararse al elaborar el marco del muestreo. Por ejemplo, realizar una muestra en dos o tres etapas reducirá los costos del muestreo y del trabajo en terreno, pero aumentarán los errores de muestreo y, por lo tanto, la precisión de las estimaciones.

Cuadro 2.2 Principales instrumentos de recopilación de datos para la evaluación de impacto

<i>Técnica</i>	<i>Definición y uso</i>	<i>Fortalezas</i>	<i>Debilidades</i>
Estudios de casos	Recopilación de información que genere un recuento que puede ser descriptivo o explicativo y puede servir para responder las preguntas cómo y por qué	<ul style="list-style-type: none"> – Pueden abordar una variedad completa de evidencias de documentos, entrevistas observación – Pueden agregar poder explicativo cuando se centran en instituciones, procesos, programas, decisiones y sucesos 	<ul style="list-style-type: none"> – Los buenos estudios de casos son difíciles de realizar – Requieren conocimientos especializados de investigación y redacción para que sean rigurosos – Los resultados no se pueden generalizar a toda la población – Tardan mucho tiempo – Son difíciles de repetir
Grupos representativos	Conversaciones focalizadas con miembros de la población beneficiaria que estén familiarizados con los temas pertinentes antes de redactar un conjunto de preguntas estructuradas. El propósito es comparar las perspectivas de los beneficiarios con conceptos abstractos de los objetivos de la evaluación.	<ul style="list-style-type: none"> – Ventajas similares a las entrevistas (a continuación) – Especialmente útiles cuando se desea una interacción entre los participantes. – Una manera útil de identificar las influencias jerárquicas 	<ul style="list-style-type: none"> – Puede ser costosos y tardar mucho tiempo – Deben ser sensible a la combinación de niveles jerárquicos – No se pueden generalizar
Entrevistas	El entrevistador plantea preguntas a una o más personas y registra las respuestas de los declarantes. Las entrevistas pueden ser formales o informales, directas o por teléfono, de interpretación cerrada o abierta.	<ul style="list-style-type: none"> – Las personas e instituciones pueden explicar sus experiencias en sus propias palabras y entorno – Son flexibles, permitiendo que el entrevistador explore vías de indagación no previstas y sondee temas en profundidad – Especialmente útiles cuando se prevén dificultades de lenguaje. 	<ul style="list-style-type: none"> – Tardan mucho tiempo – Pueden ser costosas – Si no se realizan adecuadamente, el entrevistador puede influir en la respuesta del entrevistado

Observación	Observación y registro de una situación en un registro o diario. Esto incluye quién participa; qué sucede, cuándo, dónde y cómo. La observación puede ser directa (el encuestador observa y registra) o participativa (el observador se hace parte del entorno durante un período).	<ul style="list-style-type: none"> – Mayores posibilidades de obtener información de funcionarios superiores – Proporciona información descriptiva sobre el entorno y los cambios observados 	<ul style="list-style-type: none"> – La calidad y utilidad de los datos dependen en gran medida de las capacidades de observación y redacción del encuestador – Los resultados pueden ser susceptibles a diversas interpretaciones – No se aplican fácilmente dentro de un plazo breve al cambio de proceso
Cuestionarios	Elaboración de un conjunto de preguntas de encuesta cuyas respuestas se pueden codificar coherentemente	<ul style="list-style-type: none"> – Pueden llegar simultánea a una muestra amplia – Da tiempo a los declarantes para pensar antes de responder – Se pueden responder en forma anónima – Imponen uniformidad al preguntar lo mismo a todos los declarantes – Facilita la recopilación y comparación de datos 	<ul style="list-style-type: none"> – La calidad de las respuestas depende en alto grado de la claridad de las preguntas – A veces es difícil convencer a las personas para que completen y devuelvan el cuestionario – Puede significar que hay que ajustar las actividades institucionales y las experiencias de las personas dentro de categorías predeterminadas
Análisis de documentos escritos	Revisión de documentos como registros, bases de datos administrativas, materiales de capacitación y correspondencia.	<ul style="list-style-type: none"> – Permite identificar problemas e investigarlos en más detalle y proporciona evidencia de acciones, cambios y efectos para apoyar las percepciones de los declarantes – Pueden ser costosos 	<ul style="list-style-type: none"> – Puede tardar mucho tiempo

Fuente: Adaptado de Taschereau (1998).

Una vez que el equipo de evaluación ha determinado las variables de resultado y la o las poblaciones de interés, un primer paso para el especialista en muestreo sería determinar los cálculos de poder (véase Valadez y Bamberger, 1994, págs. 382–84, para obtener un análisis del proceso de cálculo de poder). Puesto que se pueden realizar los cálculos de poder usando sólo una medida de resultados y a menudo las evaluaciones consideran varias, se deberán tomar algunas decisiones estratégicas con respecto al indicador de resultado que se usará al diseñar la muestra.

Después de elaborar la estrategia y el marco de muestreo, el especialista en muestreo también debe participar en la selección de la muestra para el trabajo en terreno y la prueba piloto, con el fin de asegurarse de que la prueba piloto no se realice en un área que se incluirá en la muestra para el trabajo en terreno. A menudo se requerirá un trabajo en terreno preliminar como parte del procedimiento de selección de la muestra. Por ejemplo, será necesario un proceso de enumeración si no existen mapas actualizados de las unidades necesarias para la muestra (hogares, escuelas, etc.) o si debe identificarse cierta población de interés, como niños desnutridos, de modo que se pueda seleccionar para el propósito de la evaluación.

Una vez finalizado el trabajo en terreno, el especialista en muestreo debe ayudar a determinar las ponderaciones de muestreo, para calcular los factores de expansión y corregir por errores de muestreo y la falta de respuestas (en Grosh y Muñoz, 1996, se presenta un análisis detallado de los procedimientos de muestreo como parte de un trabajo de encuesta domiciliaria. Kish, 1965, está considerado como uno de los libros de texto estándar en el área del muestreo). Y por último, el especialista en muestreo debe producir un documento de muestreo en que se detalle la estrategia de muestreo, incluidos (a) de la etapa de diseño de muestreo, los cálculos de poder con variables de efecto, la determinación de errores y tamaños de muestreo, el uso de estratificación para analizar poblaciones de interés; (b) de la etapa de selección de muestra, un resumen de las etapas de muestreo y procedimientos de selección; (c) de la etapa de trabajo en terreno para prepararse para el análisis, la relación entre el tamaño de la muestra y la población de la cual se seleccionó, tasas de falta de respuesta y otra información utilizada para moldear las ponderaciones del muestreo; y cualquier información adicional que el analista necesite para informar el uso de los datos de evaluación. Este documento se puede emplear para mantener los registros del proyecto de evaluación y se debe incluir junto a los datos siempre que sean distribuidos para ayudar a orientar a los analistas en utilizar los datos de la evaluación.

Cuestionarios. El diseño del cuestionario es importante para la validez de la información recopilada. Existen cuatro tipos generales de información necesaria para realizar una evaluación de impacto (Valadez y Bamberger, 1994). Estos incluyen

- Clasificación de los datos nominales y clasificación de los declarantes según si son participantes en el proyecto o si pertenecen al grupo de comparación;
- Exposición a las variables de tratamiento en que se registra no sólo los servicios y beneficios recibidos, sino también la frecuencia, cantidad y calidad; evaluar la calidad

puede ser bastante difícil;

- Variables de resultados para medir los efectos de un proyecto, incluidos los productos inmediatos, resultados sostenidos o la entrega permanente de servicios durante un período prolongado, y los efectos del proyecto (como el mejoramiento del ingreso y el empleo) y
- Variables de intervención que afectan la participación en un proyecto o el tipo de efecto producido, como características de los individuos, de los hogares o de la comunidad; estas variables pueden ser importantes para examinar los sesgos.

La forma en que se plantea la pregunta, así como el ordenamiento de las preguntas, también es bastante importante al momento de generar información confiable. Un ejemplo adecuado es la medición de bienestar, la que sería necesaria para medir el efecto directo de un proyecto en la reducción de la pobreza. El preguntar a los individuos acerca de su nivel de ingresos no necesariamente generaría resultados exactos sobre su bienestar económico. Como se analiza en la literatura sobre medición del bienestar, en general se usan preguntas sobre gastos, composición del hogar, activos, donaciones y transferencias de beneficios y el valor imputado de los alimentos cultivados en casa y de la vivienda ocupada por el propietario para captar el verdadero valor del bienestar del hogar y de los individuos. El recordatorio de tiempo usado para las partidas de gasto o el orden en que se planteen estas preguntas puede afectar significativamente la validez de la información recolectada.

Se ha observado que entre los elementos para un buen cuestionario se encuentra lo siguiente: mantener su brevedad y su concentración en preguntas importantes, asegurar que las instrucciones y las preguntas sean claras, limitar las preguntas a aquellas necesarias para la evaluación, incluida una opción “no opina” para las preguntas cerradas para garantizar datos confiables y usar procedimientos sólidos para aplicar el cuestionario, que desde luego pueden ser diferentes para las encuestas cuantitativas y las cualitativas.

Aspectos del trabajo en terreno: Trabajar con personal local con vasta experiencia en recopilación de datos similares a los que se necesitan para la evaluación puede facilitar enormemente las operaciones del trabajo en terreno. Este personal no sólo puede aportar los conocimientos requeridos sobre el territorio geográfico que se debe cubrir, sino que sus conocimientos también serán decisivos para elaborar las normas que se usarán para ubicar y abordar a los informantes. El personal en terreno cuyos conocimientos especializados corresponden a un área distinta a la requerida en el esfuerzo de evaluación puede presentar problemas, como sucedió en una evaluación de la educación en Nicaragua en que se utilizó una empresa especializada en encuestas de opinión pública para realizar una encuesta en escuelas y hogares. Los conocimientos especializados que habían permitido que esta empresa lograra una excelente reputación basada en sus predicciones precisas sobre resultados improbables en las elecciones no fueron útiles para saber la forma en que abordar a niños escolares o combinar conjuntos de datos cuantitativos. Esta falta de conocimientos especializados creó considerables problemas en la implementación de las encuestas, requiriendo semanas de medidas correctivas por parte de un equipo conjunto del Ministerio de Educación y el Banco Mundial.

El tipo de personal necesario para recolectar datos en terreno será diferente de acuerdo con los objetivos y enfoque de la evaluación. Por ejemplo, una evaluación cuantitativa de los efectos de un programa de nutrición podría requerir la inclusión de un especialista en antropometría como parte de un equipo de encuestas, para que reúna medidas de altura por peso, mientras que la evaluación de impacto de una reforma educacional probablemente incluiría personal especialista en la aplicación de pruebas de rendimiento para medir el efecto de la reforma en el rendimiento académico. La mayoría de las encuestas cuantitativas requerirá al menos un administrador de encuestas, un administrador de datos, un administrador en terreno, supervisores en terreno, entrevistadores, operadores de ingreso de datos y conductores. Según el enfoque cualitativo que se use, el personal en terreno podría ser similar, con la excepción de los operadores de ingreso de datos. Sin embargo, los conocimientos de los entrevistadores serían bastante diferentes, ya que los entrevistadores cualitativos requieren una capacitación especializada, en particular para los grupos representativos, observación directa, etc.

Otros tres puntos son útiles de recordar al planificar las operaciones de una encuesta. Primero, es importante tomar en cuenta los sucesos temporales que pueden afectar el éxito operativo del trabajo en terreno y la validez externa de los datos recopilados, como el calendario del año escolar, festivos, temporadas lluviosas, tiempos de cosecha o patrones de migración. Segundo, es determinante hacer pruebas piloto con los instrumentos de recolección de datos, aun cuando sean adaptaciones de instrumentos que se hayan usado anteriormente, para probar la calidad del instrumento con respecto a la generación de los datos requeridos y familiarizar al personal de trabajo en terreno con la dinámica del proceso de recolección de datos. Las pruebas piloto también pueden servir como ensayo para la selección de un equipo básico de personal en terreno que lleve a cabo la encuesta real. Muchos recopiladores de datos con experiencia comenzarán con un 10% a 20% más de personal en la prueba piloto de lo que se usará en el trabajo en terreno real y luego seleccionarán a los de mejor rendimiento en la prueba para que integren los equipos de recolección de datos reales. Por último, las comunicaciones son esenciales para las operaciones en terreno. Por ejemplo, si las condiciones locales permiten su uso, se puede mejorar el trabajo en terreno si los supervisores cuentan con teléfonos celulares, de manera que puedan mantenerse en contacto con el administrador de la encuesta, el administrador en terreno y otros miembros del personal para responder preguntas y mantenerlos informados sobre el progreso.

Administración datos y acceso a ellos. Un buen sistema de administración de datos debe garantizar la oportunidad y calidad de los datos de la evaluación. Su oportunidad dependerá de la máxima integración posible entre la recopilación de datos y el procesamiento, de modo que se puedan verificar y corregir errores antes de completar el trabajo en terreno. Se puede garantizar la calidad de los datos aplicando verificaciones de coherencia para probar la validez interna de los datos recolectados durante y después del ingreso de los datos y asegurándose de que los analistas que usen los datos dispongan de la documentación adecuada. La documentación debe comprender dos tipos de información: (a) la información necesaria para interpretar los datos, incluidos libros de códigos, diccionarios de datos, guías para variables creadas y todas las traducciones necesarias; y (b) la información necesaria para realizar el análisis, la que con frecuencia se incluye en un documento de información

básica que contiene una descripción del enfoque y objetivo de la evaluación, detalles sobre la metodología empleada, resúmenes o copias de los instrumentos de recolección de datos, información sobre la muestra, una descripción del trabajo en terreno y directrices para usar los datos.

Se recomienda que los datos producidos por las evaluaciones sean de disponibilidad general, dado el gran valor público de éstas y que posiblemente sea necesario que un equipo distinto al que llevó a cabo el trabajo de evaluación original realice un trabajo de seguimiento adicional para evaluar los efectos a largo plazo. Para facilitar el proceso de intercambio de datos, al principio de la evaluación se debe acordar y firmar una política de acceso abierto a los datos que establezca normas y responsabilidades para su distribución. Una política de acceso abierto a los datos agrega una carga adicional a una buena documentación de los datos y a la protección de la confidencialidad de los informantes. Si distintos organismos recopilan los datos de panel de los mismos informantes en el tiempo, éstos se deberán identificar para realizar el trabajo de seguimiento. Este requisito debe equilibrarse con las normas de confidencialidad que generalmente acompañan cualquier investigación en el sector social. Existe la posibilidad de poner los datos de registro de las unidades anónimas a disposición de todos los analistas interesados, pero solicitarle a los investigadores interesados en realizar un trabajo de seguimiento que se comuniquen con el organismo a cargo de los datos para obtener el listado de las unidades de la muestra. Esto permite que el organismo garantice el control de calidad de un trabajo futuro a través del contacto con los investigadores que pretendan realizarlo.

Análisis, información y difusión

Al igual que con otras etapas del proceso de evaluación, el análisis de los datos de la evaluación, ya sean cuantitativos o cualitativos, requiere la colaboración entre los analistas, generadores de datos y autoridades responsables para aclarar preguntas y garantizar resultados oportunos y de calidad. Seguramente surgirán problemas con la depuración e interpretación de los datos durante el análisis y se requerirán los aportes de diversos miembros del equipo.

En el Capítulo 3 se incluyen algunas de las técnicas y desafíos que plantea un análisis cuantitativo basándose en métodos estadísticos. También existen muchas técnicas para analizar datos cualitativos (véase Miles y Huberman, 1994). Aunque un análisis detallado de estos métodos escapa al horizonte de este manual, se mencionan dos métodos que se usan comúnmente para la evaluación de impacto: análisis de contenido y análisis de casos (Taschereau, 1998).

El análisis de contenido se usa para analizar datos obtenidos en entrevistas, observaciones y documentos. Al revisar los datos, el evaluador elabora un sistema de clasificación para los datos, organizando la información de acuerdo con (a) las preguntas de evaluación para las cuales se recopiló la información, (b) cómo se usará el material y (c) la necesidad de realizar referencias cruzadas con la información. La codificación de los datos puede ser bastante compleja y podría requerir varios supuestos. Una vez establecido un sistema de clasificación, comienza la fase de análisis, que también es un proceso difícil. Consiste en

buscar patrones en los datos e ir más allá de la descripción hacia la comprensión de los procesos, resultados y efectos del programa. La mejor forma de realizarlo es con la participación de los miembros del equipo. Ahora también se dispone de nuevos programas computacionales etnográficos y lingüísticos, diseñados para apoyar el análisis de datos cualitativos.

El análisis de casos se basa en estudios de casos diseñados para un estudio cabal de un determinado grupo o individuo. El alto nivel de detalles puede proporcionar valiosa información para evaluar el impacto del proyecto. Los procesos de recolectar y analizar los datos se lleva a cabo en forma simultánea, puesto que los evaluadores realizan observaciones mientras recopilan la información. Luego pueden elaborar y probar explicaciones y vincular partes cruciales de la información.

Ya sea que se trata de analizar información cuantitativa o cualitativa, también se pueden obtener algunas otras lecciones relacionadas con el análisis, información y difusión a partir de los ejemplos de casos del Anexo 1.

Primero, el análisis habitualmente demora más de lo previsto, especialmente si los datos no son tan depurados o accesibles al principio del análisis, si los analistas no tienen experiencia con el tipo de trabajo de evaluación o si se le da importancia a la creación de capacidades mediante un trabajo en colaboración. En la revisión de los estudios de casos que se consideraron para este artículo, el análisis más rápido tardó aproximadamente un año después de generar los datos y el análisis más prolongado cerca de dos años. El caso del Capítulo 3 ilustra algunos de los numerosos pasos que significa el análisis y el motivo por el cual puede tardar más de lo previsto.

Segundo, el administrador de la evaluación debe planificar la generación de diversos productos como resultados del trabajo analítico, considerando dos elementos. En primer lugar, garantizar la secuencia cronológica de los resultados en torno a sucesos clave cuando se deben tomar decisiones acerca del futuro del proyecto, como revisiones de mediados de período, elecciones o término de una fase piloto. En segundo lugar, el público para los resultados. Los productos se deben diferenciar según el público hacia los cuales están orientados, como autoridades responsables de gobierno, administradores de programa, donantes, el público en general, periodistas y académicos.

Tercero, los productos tendrán mayor aplicación en la política si incluyen recomendaciones claras y prácticas provenientes del análisis de los efectos. Éstos se pueden dividir en prioridades a corto y largo plazo y, cuando sea posible, deben incluir inferencias presupuestarias. Las autoridades directivas deben estar dispuestos a considerar los resultados finales.

Por último, los informes se deben planificar como parte de una estrategia de difusión más amplia, que puede incluir presentaciones para diversos públicos destinatarios, boletines de prensa, retroalimentación a los informantes y la publicación de la información en el Web. Esta estrategia de difusión se debe incluir en las etapas iniciales del proceso de planificación, para garantizar que se incorpore al presupuesto y que los resultados lleguen al

público previsto.