

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 aislandolo 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 seleccionando 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 $R = \text{costo}/\text{unidad (o beneficio)}$. 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 capten 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.

Capítulo 3

Aplicación de métodos analíticos para la evaluación de impacto: estudio de caso*

Este estudio de caso se basa en un programa hipotético contra la pobreza, PROSCOL, que proporciona transferencias en efectivo orientadas hacia familias pobres con hijos en edad escolar en una región de un determinado país en desarrollo. Con este caso se desean ilustrar los pasos analíticos que componen la puesta en práctica de una evaluación de impacto y las opciones que puede enfrentar un analista, en un proceso que es aplicable a cualquier tipo de programa contra la pobreza. Al examinar la forma de evaluar el impacto del programa, el analista de políticas incurre en diversos errores comunes a lo largo del proceso y busca información sobre temas específicos recurriendo a los conocimientos especializados de colegas: un estadístico, un economista, un profesor de econometría y un sociólogo.

Entre los pasos analíticos que el analista sigue en el caso están los siguientes: identificar las preguntas que se deben abordar en la evaluación del impacto, evaluar los datos disponibles, dar un primer vistazo a los datos, comprender los sesgos, aprender acerca del ingreso no percibido, agregar variables de control, comprender la importancia de la exogeneidad, explorar mejores formas de crear un grupo de comparación (correspondencia de puntuación de la propensión), conocer los sesgos debido a elementos no observables, examinar lo que se podría haber hecho con una encuesta básica o de referencia (dobles diferencias), usar variables instrumentales, probar las diversas metodologías, incorporar información recogida en el terreno y planificar trabajo futuro.

Descripción del programa hipotético, PROSCOL

En el programa PROSCOL se identifican las familias calificadas para participar usando diversas variables sustitutivas de la pobreza, entre ellas el número de personas en el hogar, la educación del jefe de hogar y diversos atributos de la vivienda. PROSCOL paga una cantidad fija por niño en edad escolar a todos los hogares seleccionados, a condición de que los niños asistan al 85% de sus clases escolares. Esto debe verificarse mediante una nota de la escuela. Los hogares deben mantener a sus hijos en la escuela hasta los 18 años.

Este programa se introdujo hace 12 meses, está financiado por el Banco Mundial y funciona a partir del Ministerio de Desarrollo Social. En un esfuerzo por evaluar el impacto de PROSCOL en la pobreza y así ayudar a determinar si se debe ampliar el programa hasta abarcar el resto del país o abandonarse, el Banco Mundial ha solicitado que el Ministerio de Finanzas evalúe sus efectos. La solicitud se hizo al Ministerio de Finanzas, para así asegurar una evaluación independiente y para ayudar a crear capacidades para este tipo de evaluaciones en una unidad central del gobierno, una unidad cercana a la entidad en que se

* Este capítulo se basa en gran medida en un documento de información básica de Martin Ravallion, *The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation*, Documento de trabajo de estudio de políticas N° 2153, 1999.

realizan las asignaciones presupuestarias.

Identificación de las preguntas que se abordarán en la evaluación del impacto

El primer paso para el analista del Ministerio de Finanzas asignado a la tarea de la evaluación de PROSCOL es aclarar los objetivos del proyecto que se deberán considerar al evaluar sus impactos. El proyecto tiene dos objetivos: uno, las transferencias de efectivo, destinadas a reducir la pobreza actual y dos, al insistir en que los destinatarios de la transferencia mantengan a sus niños en la escuela, intentar reducir la futura pobreza aumentando los niveles de educación entre la actual población de niños pobres. Por consiguiente, se necesitarán dos elementos de información sobre el programa para evaluar sus impactos. Primero ¿las transferencias de efectivo llegan principalmente a familias de bajos ingresos? Y segundo, ¿en qué medida el programa está aumentando las tasas de matrícula escolar?

Evaluación de los datos disponibles

Para realizar la evaluación, el analista cuenta con dos recursos principales. El primero es un informe basado en entrevistas cualitativas a los administradores de programas y grupos representativos de participantes. Sin embargo, no está claro si los entrevistados eran representativos de los participantes en PROSCOL o cuán pobres eran con respecto a los que no se seleccionaron para el programa y no fueron entrevistados. El informe señala que los niños asistían a la escuela, pero es posible que también habrían ido a la escuela si no hubiera existido el programa. Aunque este informe es un punto de partida importante, no indica al analista el nivel de pobreza de los participantes en PROSCOL y cómo el programa afecta la escolaridad. La segunda fuente es una encuesta domiciliaria independiente a nivel nacional realizada por la Oficina de Estadísticas del país, llamada Encuesta Sobre las Condiciones de Vida (ENCOVI). La ENCOVI incluía una muestra aleatoria de 10.000 hogares y consultaba acerca de los ingresos del hogar por fuente, empleo, gastos, estado de salud, grados de instrucción, atributos demográficos y otras características de la familia. La encuesta incluía una pregunta sobre si el hogar muestreado había participado en PROSCOL y una partida para el dinero recibido de PROSCOL en el listado de las fuentes de ingresos.

Un primer examen de los datos

Luego el analista se consigue los datos sin elaborar de la ENCOVI, con el fin de evaluar quién se beneficia con el programa. Utiliza un paquete de software estadístico, como SPSS o SAS, para generar una tabulación cruzada de la cantidad promedio recibida de PROSCOL por los deciles de hogares, donde los deciles se crean clasificando todos los hogares de la muestra según su ingreso por persona. Para calcular esto último, el analista decide restar todos los dineros recibidos de PROSCOL como una buena medida de los ingresos en ausencia del programa, y así identificar quién se benefició con respecto a su ingreso anterior a la intervención.

La tabulación cruzada sugiere que las transferencias de efectivo del programa están orientadas adecuadamente hacia los pobres. Según la línea de pobreza oficial del país, aproximadamente el 30% de la población del noroeste es pobre. Los cálculos del cuadro indican que el 30% más pobre de la muestra de la encuesta recibe el 70% de las transferencias de PROSCOL. A primera vista, esto parece ser un resultado positivo.

La pregunta siguiente es acerca del impacto en la escolaridad. Éste se examina a través de una tabulación cruzada de las tasas de matrícula escolar promedio de diversos grupos de edades para familias dentro de PROSCOL versus familias fuera de PROSCOL. La tabulación casi no muestra diferencia alguna entre ambos grupos; la tasa de matrícula promedio para niños entre 6 y 18 años es de alrededor del 80% en los dos casos. Luego el analista calcula los años promedio de escolaridad para cada edad y los resultados se representan por separado para las familias dentro de PROSCOL y las familias fuera de éste. Si bien las dos cifras no son idénticas, son muy cercanas. En esta etapa, el analista se pregunta si en realidad el programa no ha influido en la escolaridad o si el enfoque es incorrecto.

Comprensión de los sesgos

Con esta incertidumbre, el analista luego busca información de un estadístico superior para que averigüe por qué los resultados sugieren que los niños del PROSCOL no tienen más posibilidades de estar en la escuela que los niños fuera del PROSCOL. El estadístico plantea la hipótesis de que los resultados pueden tener un sesgo importante. Para evaluar el impacto del programa, es necesario saber qué habría sucedido sin el programa. Sin embargo, el analista no ha tomado en consideración este punto; en su lugar, las familias fuera del PROSCOL se usan como grupo de comparación para inferir lo que habría sucedido con la escolaridad de los participantes del PROSCOL si no hubiera existido el programa.

En otras palabras, P_i representa la participación en PROSCOL del i -ésimo niño. Esto puede tener dos valores posibles, a saber, $P_i = 1$ si el niño participa en PROSCOL y $P_i = 0$ si no lo hace. Si el i -ésimo niño no participa, su nivel de escolaridad es S_{0i} , lo que representa la escolaridad S del niño i cuando $P = 0$. Si el niño participa, su escolaridad es S_{1i} . Su ganancia en la escolaridad debido a PROSCOL es $S_{1i} - S_{0i}$. Luego, la ganancia para el i -ésimo niño que participe ($P = 1$) es

$$G_i = S_{1i} - S_{0i} \mid P_i = 1.$$

El símbolo \mid significa “dado que” o “bajo la condición de que” y es necesario para dejar en claro que el cálculo es la ganancia para un niño que realmente participó. Si se desea saber la ganancia promedio, ésta simplemente es la media de todas las G_i , con lo que se obtiene la ganancia media en escolaridad entre todos los que participaron en PROSCOL. Siempre que esta media se calcule en forma correcta (usando las ponderaciones de muestras adecuadas de la encuesta), proporcionará una estimación insesgada de la verdadera ganancia media. Esta última es el “valor esperado” de G y puede representarse como

$$G = E(S_{Ii} - S_{0i} \mid P_i = 1).$$

Esta es otra forma de decir “media”. Sin embargo, no necesita ser exactamente igual a la media calculada de los datos de la muestra, dado que habrá algún error de muestreo. En la literatura sobre evaluación, $E(S_{Ii} - S_{0i} \mid P_i = 1)$ a veces se denomina el “efecto tratamiento” o “efecto tratamiento promedio de los tratados”. En este caso, PROSCOL es el tratamiento.

El estadístico señala al analista que éste no ha calculado G , sino más bien la diferencia en la escolaridad media entre los niños de familias beneficiadas por PROSCOL y los de familias fuera de PROSCOL. Esta es la estimación de muestra de

$$D = E(S_{Ii} \mid P = 1) - E(S_{0i} \mid P = 0)$$

Hay una simple identidad que vincula D y G , a saber:

$$D = G + B.$$

Este término “ B ” es el sesgo de la estimación y está dado por

$$B = E(S_{0i} \mid P_i = 1) - E(S_{0i} \mid P_i = 0)$$

En otras palabras, el sesgo es la diferencia esperada en escolaridad sin PROSCOL entre los niños que de hecho participaron en el programa y aquellos que no lo hicieron. Se podría corregir este sesgo si se conociera $E(S_{0i} \mid P_i = 1)$, pero ni siquiera es posible obtener una estimación de muestra de éste. No se puede observar cómo habría sido la escolaridad de los niños que realmente participaron en PROSCOL si no hubieran participado; estos son datos faltantes, también denominados una media “contrafactual”.

Este sesgo es muy importante. A falta del programa, los padres de PROSCOL probablemente habrían enviado a sus hijos a la escuela en menor proporción que otros padres. Si hubiera sido así, habrá un sesgo en el cálculo. Volviendo a las preguntas de evaluación originales, lo que interesa es la escolaridad adicional debido a PROSCOL. Supuestamente esto sólo afecta a las familias que efectivamente participan. Es decir, es necesario saber cuánta escolaridad menos se podría haber esperado sin el programa. Si no hay sesgo, la escolaridad adicional generada por el programa es la diferencia en la escolaridad media entre los que participaron y los que no lo hicieron. Así, el sesgo se genera si hay una diferencia en la escolaridad media entre los padres dentro de PROSCOL y los padres fuera de PROSCOL en ausencia del programa.

Para eliminar el sesgo, el mejor método sería asignar el programa en forma aleatoria. Entonces los participantes y no participantes tendrían la misma escolaridad esperada sin el programa, esto es, $E(S_{0i} \mid P_i = 1) = E(S_{0i} \mid P_i = 0)$. La escolaridad de las familias no participantes, por consiguiente, revelaría en forma correcta el escenario contrafactual, es decir, la escolaridad que se habría observado para los participantes si no hubieran tenido acceso al programa. Por cierto, la asignación aleatoria será igual a la distribución total, no sólo a la media. Aún existiría un sesgo debido al error de muestreo, pero para muestras lo suficientemente grandes se puede suponer con seguridad que cualquier diferencia

estadística significativa en la distribución de la escolaridad entre los participantes y no participantes es atribuible al programa.

Dentro del actual diseño del programa, es claro que la participación no es aleatoria. Por cierto, sería una seria crítica para PROSCOL descubrir que sí lo era. El mismo hecho de su orientación intencional hacia las familias pobres (que supuestamente tienen menos posibilidad de enviar a sus hijos a la escuela) crearía un sesgo.

Esto plantea la siguiente pregunta: si PROSCOL funciona bien, se debería esperar que los participantes tuvieran peor escolaridad sin el proceso. Luego $E(S_{0i} | P_i = 1) < E(S_{0i} | P_i = 0)$. y el cálculo original de los analistas subestimaría la ganancia del programa. Se podrían encontrar muy pocas o ninguna ganancia, aun cuando el programa estuviera funcionando en forma adecuada.

Ahora el analista examina si la magnitud de este sesgo podría ser considerable. Supóngase que las familias pobres envían a sus hijos a trabajar en lugar de la escuela; como son pobres y no pueden endeudarse fácilmente, necesitan dinero adicional ahora. Las familias que no son pobres envían a sus niños a la escuela. El programa selecciona a familias pobres, las que entonces envían a sus niños a la escuela. Se observa una diferencia insignificante en la escolaridad media entre las familias de PROSCOL y las familias fuera de éste; en efecto, $E(S_{1i} | P_i = 1) = E(S_{0i} | P_i = 0)$ en valor esperado. Pero el impacto del programa es positivo y está dado por $E(S_{0i} | P_i = 0) - E(S_{0i} | P_i = 1)$. El no haber considerado la focalización intencional del programa hacia los pobres, bien podría haber producido una subestimación considerable de los beneficios de PROSCOL cuando el analista comparó la escolaridad media entre familias dentro de PROSCOL y familias fuera de éste.

Adquisición de conocimientos sobre el ingreso no percibido

A continuación, el analista muestra a otro colega, un economista del Ministerio de Finanzas, los resultados de su tabulación cruzada de cantidades recibidas de PROSCOL versus el ingreso. El economista plantea un aspecto importante: las ganancias que los pobres obtuvieron de PROSCOL claramente fueron sobreestimados, porque se ha ignorado el ingreso no percibido o sacrificado. Los niños deben ir a la escuela si la familia desea recibir la transferencia de PROSCOL; por consiguiente no podrán trabajar en el negocio de la familia ni en el mercado laboral. Por ejemplo, los niños entre 15 y 18 años pueden ganar dos tercios o más del salario de un adulto en el sector agrícola y de la construcción. Las familias que participan en PROSCOL perderán este ingreso proveniente del trabajo de sus hijos. Este ingreso no percibido se debe tomar en cuenta al calcular las ganancias de ingreso netas que produce el programa. Y se debe restar esta ganancia de ingreso neta, no la transferencia bruta, para calcular el ingreso previo a la intervención. Esto también tendrá importancia al determinar qué tan pobre habría sido la familia en ausencia de la transferencia de PROSCOL. En consecuencia, el cuadro actual podría sobrestimar en forma considerable los beneficios que tiene el programa para los pobres.

El analista se pregunta por qué debe descontar el ingreso no percibido del trabajo de los niños, dado que menos trabajo infantil es algo positivo. El economista explica que debe examinar los beneficios de un menor trabajo infantil, del cual el principal beneficio es la

escolaridad adicional y, por ende, futuros ingresos más altos para las familias actualmente pobres. El analista ha generado cuadros que reflejan las dos formas principales en que PROSCOL reduce la pobreza: aumentando los actuales ingresos de los pobres y aumentando sus ingresos futuros. El impacto en el trabajo infantil es importante para ambos casos, pero en sentidos opuestos; por consiguiente, PROSCOL representa una relación inversa de compensación.

Esto pone de relieve por qué es importante obtener una buena estimación del impacto en la escolaridad; sólo entonces será posible determinar el ingreso no percibido. Por ejemplo, es posible que el tiempo adicional en la escuela sea en desmedro del tiempo no trabajado.

Con respecto a la segunda tabulación cruzada, el aspecto principal planteado por el economista es que no se consideran todas las demás determinantes de la escolaridad, aparte de la participación en PROSCOL. Él sugiere ejecutar una regresión de los años de escolaridad en un conjunto de variables de control y también si la familia del niño estaba o no cubierta por PROSCOL. Para el i -ésimo niño de la muestra, sea

$$S_i = a + bP_i + cX_i + \varepsilon_i .$$

Aquí a , b y c son parámetros; X representa las variables de control, como edad del niño, educación del padre y de la madre, tamaño y composición demográfica del hogar y características de la escuela; y ε es un valor residual que incluye otras determinantes de la escolaridad y errores de medición. El valor estimado de b proporciona el impacto de PROSCOL en la escolaridad.

Cabe observar que si la familia del i -ésimo niño participa en PROSCOL, entonces $P = 1$ y su escolaridad será $a + b + cX_i + \varepsilon_i$. Si no participa, luego $P = 0$ y su escolaridad será $a + cX_i + \varepsilon_i$. La diferencia entre ambos es la ganancia en la escolaridad debido al programa, que es exactamente b .

Incorporación de variables de control

Como se sugirió, el analista luego ejecuta una regresión con y sin las variables de control. Cuando la ejecuta sin ellas, los resultados muestran que el valor estimado de b no es significativamente distinto de cero (usando la prueba t estándar que proporciona el paquete estadístico). Estos resultados parecen ser muy similares a los resultados finales, considerando la diferencia en las medias entre los participantes y los no participantes, lo que sugiere que PROSCOL no está produciendo ningún impacto en la escolaridad. Sin embargo, cuando se incluyen diversas variables de control en la regresión, hay un coeficiente positivo y significativo en la participación en PROSCOL. El cálculo muestra que a la edad de 18 años, el programa ha agregado dos años a la escolaridad.

El analista se pregunta, ¿por qué estas variables de control afectan tanto? Además, ¿se están usando las variables de control correctas? Luego visita a su ex profesor de econometría y le muestra sus regresiones. El primer aspecto relacionado con la regresión de escolaridad en P y X que plantea el profesor es que ésta no permite que el impacto del programa varíe con X ;

el impacto es igual para todo el mundo, lo que no parece ser muy probable. Es más probable que los padres con mayor escolaridad envíen a sus hijos a la escuela, de modo que para ellos los beneficios de PROSCOL serán menores. Para permitir que los beneficios varíen con X , sea la escolaridad media de los no participantes $a_0 + c_0X_i$ y la de los participantes, $a_1 + c_1X_i$, de modo que el nivel observado de escolaridad es

$$S_i = (a_1 + c_1X_i + \varepsilon_{1i})P_i + (a_0 + c_0X_i + \varepsilon_{0i})(1 - P_i),$$

donde ε_0 y ε_1 son errores aleatorios, cada uno con medias de cero y no correlacionados con X . Para estimar este modelo, a la regresión ya ejecutada se debe agregar un término adicional para los efectos de interacción entre la participación en el programa y las características observadas. Por lo tanto, la regresión ampliada es

$$S_i = a_0 + (a_1 - a_0)P_i + c_0X_i + (c_1 - c_0)P_i X_i + \varepsilon_i,$$

donde $\varepsilon_i = \varepsilon_{1i}P_i + \varepsilon_{0i}(1 - P_i)$. Luego $(a_1 - a_0) + (c_1 - c_0)X$ es el impacto medio del programa en cualquier valor dado de X . Si se usa la media X en la muestra de participantes, ésta arrojará la ganancia media del programa.

Comprensión de la importancia de la exogeneidad

Otro punto que planteó el profesor de econometría es la forma en que se estimó la regresión. Al usar el comando de regresión en el paquete estadístico, los mínimos cuadrados ordinarios (MCO), hay un problema porque las estimaciones de MCO de los parámetros estarán sesgadas incluso en muestras grandes, a menos que las variables del lado derecho sean exógenas. Exogeneidad significa que las variables del lado derecho estén determinadas en forma independiente de las opciones de escolaridad y que de este modo no estén correlacionadas con el término de error en la regresión de escolaridad. Dado que la participación en el programa se orientó intencionalmente hacia grupos específicos, la participación en PROSCOL no es exógena. Esto puede afectar el cálculo del impacto del programa de la siguiente forma: La ecuación de los años de escolaridad es

$$S_i = a + bP_i + cX_i + \varepsilon_i.$$

Se empleó el valor de $a + b + cX_i + \varepsilon_i$ como la estimación de la escolaridad del i -ésimo hogar cuando participa en PROSCOL, mientras que se usó $a + cX_i + \varepsilon_i$ para estimar la escolaridad si éste no participa. Así, la diferencia, b , es la ganancia que arroja el programa. Sin embargo, al efectuar este cálculo, el supuesto implícito es que ε_i era igual en cualquier caso. En otras palabras, el supuesto era que ε_i era independiente de P , lo que afectaría el cálculo del impacto del programa.

Esto destaca el sesgo que genera la asignación no aleatoria del programa, lo que también podría estar afectando la estimación basada en el modelo de regresión sugerido anteriormente por el economista ($S_i = a + bP_i + cX_i + \varepsilon_i$). Sin embargo, esto no significa que los resultados sean completamente incorrectos.

El profesor de econometría aclara esto con una ecuación explícita para P , a saber,

$$P_i = d + eZ_i + v_i,$$

donde Z representa diversas variables que incluyen todos los “sustitutos de pobreza” usados para la focalización de PROSCOL. También habrá algún término de error puramente aleatorio que influye en la participación; éstos son sustitutos de la pobreza no representados en los datos; y también habrá errores en la selección de participantes que se incluyen en este término v . Esta ecuación es lineal, aunque P sólo puede tomar dos valores posibles, 0 y 1. Los valores predichos entre cero y uno son aceptables, pero un modelo lineal no puede eliminar la posibilidad de valores predichos negativos o de valores superiores a uno. Existen modelos no lineales que pueden tratar este problema, pero para simplificar el análisis, será más fácil confinar la atención en los modelos lineales.

Existe un caso especial en que la anterior regresión MCO de S sobre P y X dará una estimación insesgada de b . Esto es cuando X incluye todas las variables en Z que también influyen en la escolaridad, y el término de error v no está correlacionado con el término de error ε en la regresión para la escolaridad. En la literatura sobre evaluación, esto a veces se denomina “selección sobre observables”.

Supóngase que las variables de control X en la regresión anterior para la escolaridad incluyen todas las variables observadas Z que influyen en la participación P y v no está correlacionado con ε (de modo que las variables no observadas que afectan la asignación del programa no influyen en la escolaridad, dependiendo de X). Luego esto ha eliminado cualquier posibilidad de que P esté correlacionada con ε . Ésta ahora será exógena en la regresión para la escolaridad. En otras palabras, la idea clave de la selección sobre observables es que hay algún X observable, de modo que el sesgo desaparece dependiendo de X .

La adición de las variables de control a la regresión de escolaridad sobre la participación en PROSCOL produjo una gran diferencia, porque X debe incluir variables que estaban entre los sustitutos de pobreza usados para la focalización, o que estaban correlacionados con éstos, y son variables que también influyen en la escolaridad. Sin embargo, esto sólo funciona si los supuestos son válidos. Se deben tener en cuenta dos problemas. En primer lugar, el método anterior falla si no hay determinantes no observadas de la participación, es decir, si el término de error v tienen varianzas cero y todas las determinantes de la participación también afectan la escolaridad. Entonces no hay una variación independiente en la participación en el programa que permita identificar su impacto en la escolaridad; es perfectamente posible predecir P a partir de X y entonces no estimará la regresión. Es poco probable que este problema surja a menudo, dado que casi siempre existen determinantes no observadas de la asignación de un programa.

El segundo problema es más común y en este caso, más preocupante. El término de error ε en la regresión de la escolaridad probablemente contiene variables que no se encuentran en la encuesta, pero bien podrían influir en la participación en el programa, es decir, podrían estar correlacionadas con el término v en la ecuación de participación. Si esto sucede, entonces $E(\varepsilon|X,P) \neq 0$ y los métodos de regresión ordinaria todavía estarán sesgados cuando

se estimen las regresiones para la escolaridad. Así, el punto clave es la magnitud de la correlación entre el término de error de la ecuación para la participación y el correspondiente a la ecuación de la escolaridad.

Búsqueda de mejores formas para crear un grupo de comparación: correspondencia de puntuación de la propensión

Contando con más información del profesor, el analista aprende que existen mejores formas de crear un grupo de comparación. El objetivo es comparar los niveles de escolaridad según las características observadas. Si los grupos de la muestra se dividen en grupos de familias con valores iguales o similares de X , se comparará la media condicional para las familias dentro de PROSCOL y las familias fuera de PROSCOL. Si, en ausencia del programa, la escolaridad es independiente de la participación, dado X , la comparación generará una estimación insesgada del impacto de PROSCOL. Esto a veces se denomina “independencia condicional” y es el supuesto clave formulado por todos los métodos de grupos de comparación.

Así, una mejor forma de seleccionar un grupo de comparación, dados los datos existentes, es usar, para cada participante, un no participante con las mismas características observadas como control. Sin embargo, esto podría resultar muy difícil, porque el conjunto de datos podría tener muchas de esas variables. Podría no haber nadie entre los no participantes con exactamente los mismos valores de todas las características observadas para cada uno de los participantes en PROSCOL.

Un enfoque estadístico de correspondencia de puntuación de la propensión proporciona técnicas para simplificar enormemente el problema. En lugar de intentar asegurar que el control pareado para cada participante tenga exactamente el mismo valor de X , se puede lograr el mismo resultado realizando una correspondencia sobre el valor predicho de P , dado X , que se denomina la puntuación de la propensión de X . Rosenbaum y Rubin (1983) demuestran que si (en este caso) la escolaridad sin PROSCOL es independiente de la participación dado X , los participantes también son independientes de la participación dada la puntuación de la propensión de X . Puesto que la puntuación de la propensión es sólo un número, es mucho más fácil controlarla que X , que podría ser diversas variables. Y sin embargo, la correspondencia de puntuación de la propensión es suficiente como para eliminar el sesgo, siempre que haya independencia condicional dado X .

En otras palabras, primero se ejecuta una regresión de P sobre X para obtener el valor predicho de P para cada valor posible de X , que luego se estima para toda la muestra. Para cada participante, se debe encontrar el no participante con el valor más cercano a esta probabilidad predicha. Luego la diferencia en la escolaridad es la ganancia estimada del programa para ese participante.

Se puede sacar la media de todas estas diferencias para estimar el impacto. O bien, sacar la media para diferentes grupos de ingreso. Sin embargo, la forma en que se estima el modelo de la participación requiere cierta prudencia. Un modelo lineal podría proporcionar probabilidades predichas irregulares, mayor a uno, o negativas. Es mejor usar el comando

LOGIT del paquete estadístico. Este supone que el término de error v de la ecuación de participación tiene una distribución logística y estima los parámetros en forma consecutiva con el supuesto de los métodos de probabilidad máxima. Esto se basa en los principios de la estimación de probabilidad máxima de los modelos de respuesta binaria.

Otro aspecto que hay que tener en cuenta es que desde el principio se deben excluir algunos de los no participantes como posibles pares. En efecto, algunos resultados recientes dentro de la literatura sobre econometría indican que una fuente importante de sesgo en las evaluaciones es el no comparar los participantes y los controles a valores comunes de variables de pareo (véase Heckman y otros, 1998).

La intuición indica que el grupo de comparación debe ser lo más similar posible al grupo de tratamiento en términos de los elementos observables, como lo resume la puntuación de la propensión. Se podría concluir que parte de la muestra de no participantes tiene una puntuación de la propensión más baja que cualquiera de los correspondientes a la muestra de tratamiento. Esto a veces se denomina “falta de soporte común”. Al crear el grupo de comparación, se deben eliminar esas observaciones del conjunto de no participantes, para garantizar que sólo se compararán las ganancias dentro del mismo margen de puntuaciones de la propensión. También se deben excluir los no participantes para los cuales la probabilidad de participar es cero. Se recomienda reducir una pequeña proporción de la muestra, digamos un 2%, del extremo superior e inferior de la distribución de no participantes en términos de puntuaciones de la propensión. Una vez identificados los participantes y asimismo los no participantes en una región de correspondencia común, se recomienda sacar un promedio de (por ejemplo) los cinco o más cercanos en términos de diferencia absoluta en puntuaciones de la propensión (Recuadro 3.1).

Recuadro 3.1 Pasos en la correspondencia de puntuación de la propensión

El objetivo de la correspondencia es encontrar el grupo de comparación más cercano de una muestra de no participantes a la muestra de participantes en el programa. Lo “más cercano” se mide en términos de características observables. Si sólo hay una o dos de esas características, la correspondencia será más fácil. Pero es habitual que existan muchas eventuales características. Los principales pasos para la correspondencia basada en las puntuaciones de la propensión son las siguientes:

Paso 1: Se debe contar con una encuesta de una muestra representativa de no participantes calificados y una para los participantes. Mientras mayor sea la muestra de no participantes calificados, mejor, puesto que facilitará una buena correspondencia. Si las dos muestras provienen de diferentes encuestas, deberían ser encuestas muy comparables (mismo cuestionario, mismos entrevistadores o capacitación de los entrevistadores, mismo período de encuesta, etc.).

Paso 2: Reunir las dos muestras y estimar un modelo logit de la participación en el programa como función de todas las variables en los datos que probablemente determinen la participación.

Paso 3: Crear los valores predichos de la probabilidad de participar a partir de la regresión logit; éstos se denominan “puntuaciones de la propensión”. Se obtendrá una puntuación de la propensión para cada participante y no participante de la muestra.

Paso 4: Es posible que desde el principio se deba excluir a algunos de la muestra de no participantes, por tener una puntuación de la propensión fuera del margen (en general demasiado baja) con respecto al de la muestra de tratamiento. El margen de puntuaciones de la propensión estimado para el grupo de tratamiento debe ser muy parecido al de la submuestra conservada para los no participantes. Se podría tener que restringir las posibles correspondencias en otras formas, según el escenario. Por ejemplo, es posible que se decida permitir sólo correspondencias con la misma área geográfica, para asegurar que las correspondencias provengan del mismo entorno económico.

Paso 5: Para cada individuo de la muestra de tratamiento, ahora se necesita encontrar la observación en la muestra de no participantes que se acerque más en su puntuación de la propensión, medida por la diferencia absoluta en puntuaciones. Esto se denomina el “vecino más cercano”. Se pueden encontrar, por ejemplo, los cinco vecinos más cercanos.

Paso 6: Calcular el valor de la media del indicador de resultados (o cada uno de los indicadores, si hay más de uno) para los cinco vecinos más cercanos. La diferencia entre esa media y el valor real para la observación tratada es la estimación de la ganancia debido al programa para esa observación.

Paso 7: Calcular la media de las ganancias de estos individuos para obtener la ganancia promedio total. Esto se puede estratificar por medio de alguna variable de interés, como el ingreso, en la muestra de no participantes.

Esta es la forma más simple de correspondencia de puntuación de la propensión. En la práctica pueden surgir complicaciones. Por ejemplo, si se realiza un segundo muestreo de los participantes, se pueden emplear métodos de muestreo basados en la selección para corregir por esto (Manski y Lerman, 1977); como alternativa, se puede utilizar la razón de disparidad ($p/(1 - p)$, donde p es la puntuación de la propensión) para la correspondencia. En lugar de depender del vecino más cercano, se podrían utilizar todos los no participantes como posibles correspondencias, pero ponderarlos de manera diferente, según su cercanía (Heckman y otros, 1998).

Luego se deben incluir todas las variables del conjunto de datos que son o podrían ser sustitutos de los indicadores de pobreza utilizados al seleccionar los participantes en PROSCOL. Nuevamente, X debe incluir las variables en Z . Sin embargo, esto pone en evidencia una debilidad de la correspondencia de puntuación de la propensión. Con la correspondencia, un X diferente generará una estimación diferente del impacto. Con la aleatorización, el experimento ideal, los resultados no dependen de las variables X que se seleccionen. La aleatorización tampoco requiere especificar un modelo para la participación, ya sea un modelo logit u otro. En el Recuadro 3.1 se resumen los pasos para realizar la correspondencia de puntuación de la propensión.

Adquisición de conocimientos sobre sesgos debido a elementos no observables

Incluso después de formar el grupo de comparación, el analista no puede estar seguro de que esto le brinde una estimación mucho mejor del impacto del programa. Los métodos antes descritos sólo eliminarán el sesgo si existe independencia condicional, de modo que las determinantes no observables de la escolaridad (no incluidas en el conjunto de variables de control X) no estén correlacionadas con la asignación del programa. Hay dos fuentes distintas de sesgo, aquella causada por diferencias en los elementos observables y aquella causada por diferencias en los elementos no observables; estas últimas a menudo se denominan “sesgo de selección”. En el Recuadro 3.2 se explica en detalle esta diferencia.

Regresando a la última ecuación del profesor, ésta muestra que se mantendrá la independencia condicional si P es exógeno, para que entonces $E(\varepsilon_i | X_i, P_i) = 0$. Sin embargo, la asignación endógena del programa debido a la focalización intencional basada en elementos no observables todavía dejará un sesgo. Esto a veces se denomina selección sobre observables. Así, las condiciones requeridas para justificar el método planteado anteriormente por el economista no son menos restrictivas que las necesarias para justificar una versión del primer método basado en comparar familias dentro de PROSCOL con familias fuera de éste para hogares con valores similares de X . Ambos se basan en la convicción de que estos elementos no observables no influyen en conjunto en la escolaridad y la participación el programa, dado X .

En forma intuitiva, se podría pensar que una cuidadosa correspondencia reduce el sesgo, pero no necesariamente es así. La correspondencia elimina parte del sesgo en la primera estimación simplista del impacto de PROSCOL. Pero se mantiene el sesgo generado por cualquier elemento no observable difícil. Sin embargo, estas dos fuentes de sesgo se podrían compensar: una podría ser positiva y la otra negativa. Heckman y otros (1998) hacen ver esto. Así, la estimación de correspondencia bien podría tener más sesgo que la estimación simplista. No se puede saber a priori qué tanto favorece incluso un grupo de comparación bien elegido, ya que se trata de un aspecto empírico.

Revisión de lo que se podría haber hecho con una encuesta básica: Estimaciones de doble diferencia

Luego el analista consulta si habría otro método aparte de la aleatorización que sea sólido para estos elementos no observables dificultosos. Esto requeriría datos básicos o de referencia para los participantes y los no participantes recopilados antes de que se implementara el programa PROSCOL. La idea es recopilar datos sobre los resultados y sus determinantes antes y después de introducir el programa, junto con datos de un grupo de comparación sin tratamiento y del grupo de tratamiento. Luego basta con sólo restar la diferencia entre la escolaridad de los participantes y el grupo de comparación antes de introducir el programa, de la diferencia después del programa.

Recuadro 3.2 Fuentes de sesgo en estimaciones simplistas del impacto de PROSCOL

El sesgo descrito por el estadístico es la diferencia esperada en la escolaridad sin PROSCOL entre las familias seleccionadas para el programa y aquellas que no fueron elegidas. Se puede desglosar en dos fuentes de sesgo:

- Sesgo debido a diferencias en las características observables. Esto se puede originar de dos formas. En primer lugar, podría no haber soporte común. El “soporte” es el conjunto de valores de las variables de control para las que se observan los resultados y la participación en el programa. Si el soporte difiere entre la muestra de tratamiento y el grupo de comparación, esto sesgará los resultados. En efecto, no se están comparando elementos semejantes. En segundo lugar, incluso con un soporte común, la distribución de las características observables podría ser diferente dentro del área de soporte común; efectivamente, los datos del grupo de comparación están ponderados en forma incorrecta. Una selección cuidadosa del grupo de comparación puede eliminar esta fuente de sesgo.
- Sesgo debido a diferencias en las características no observables. El término sesgo de selección a veces se limita únicamente a este componente (aunque algunos autores usan ese término para el sesgo total en una evaluación no experimental). Esta fuente de sesgo surge cuando, para valores dados de X , existe una relación sistemática entre la participación en el programa y los resultados en ausencia del programa. En otras palabras, hay variables no observables que en conjunto influyen en la escolaridad y la participación en el programa dadas las variables observadas en los datos.

No hay nada que garantice que estas dos fuentes de sesgo funcionarán en la misma dirección. De modo que eliminar alguna de ellas no significa que se reducirá el sesgo total en el valor absoluto. Esa es una pregunta empírica. En uno de los pocos estudios que abordan esta pregunta, el verdadero impacto, medido por un experimento bien diseñado, se comparó con diversas estimaciones no experimentales (Heckman y otros, 1998). El sesgo en la estimación simplista fue enorme, pero una cuidadosa correspondencia del grupo de comparación basada en los observables redujo en forma considerable el sesgo.

Esto se llama la estimación de “doble diferencia” o “diferencia en las diferencias”. Abordará las variables no observadas dificultosas, siempre que no varíen en el tiempo.

Se puede explicar agregando subíndices a la ecuación anterior, de modo que se introduce la escolaridad después del programa:

$$S_{ia} = a + bP_i + cX_{ia} + \varepsilon_{ia} .$$

Antes del programa, en la encuesta básica, el rendimiento escolar en su lugar es

$$S_{ib} = a + c X_{ib} + \varepsilon_{ib} .$$

(Por cierto, $P = 0$ antes de introducir el programa). Los términos de error incluyen un impacto adicional constante en el tiempo, de modo que se pueden representar como

$$\varepsilon_{it} = \eta_i + \mu_{it} \text{ (para } t = a, b),$$

donde η_i es el impacto constante en el tiempo, que se permite que esté correlacionado con P_i , y μ_{it} es un error de innovación, que no está correlacionado con P_i (o X_i).

Aquí la idea esencial es usar los datos básicos para revelar los elementos no observables problemáticos. Cabe notar que puesto que la encuesta básica corresponde a los mismos hogares que se tienen ahora, el i -ésimo hogar de la ecuación para S_{ia} es el mismo que el i -ésimo hogar de la ecuación para S_{ib} . Luego se puede sacar la diferencia entre la ecuación “después” y la ecuación “antes”:

$$S_{ia} - S_{ib} = bP_i + c(X_{ia} - X_{ib}) + \mu_{ia} - \mu_{ib}$$

Ahora es posible hacer una regresión del cambio en la escolaridad y la participación en el programa y los cambios en X . MCO proporcionará una estimación insesgada del impacto del programa. Se han eliminado los elementos no observables, aquellos que están correlacionados con la participación en el programa.

Dado esto, si la asignación del programa se basara sólo en variables (tanto observadas como no observadas) que se conocieran en el momento de la encuesta básica, sería razonable suponer que los η no cambian entre las dos encuestas. Esto se mantendría siempre que los no observables problemáticos fueran constantes en el tiempo. Los cambios en la escolaridad en el tiempo para el grupo de comparación revelarán lo que habría sucedido al grupo de tratamiento sin el programa.

Para ello se necesitaría conocer bien el programa y poder sincronizar las encuestas de evaluación, de modo de coordinarlas con el programa. De lo contrario, inevitablemente serían los cambios no observados después de la encuesta básica los que influyen en quién se ve favorecido con el programa. Esto crearía ceros que cambiarían entre las dos encuestas.

Esta última ecuación se puede interpretar como que las características del niño y del hogar en X no son pertinentes para el cambio en la escolaridad si estas características no varían en el tiempo. Pero el beneficio de la escolaridad podría depender de la educación de los padres (y no sólo de cualquier cambio en su educación) y posiblemente del lugar donde se encuentra el hogar, porque esto determinará el acceso a las escuelas. También puede haber situaciones en las cuales los cambios en el tiempo en el indicador de resultados están influidos por las condiciones iniciales. Luego sólo se necesitará controlar por las diferencias en las condiciones iniciales. Esto se puede hacer agregando simplemente X_a y X_b por separado a la regresión, de modo que la regresión adopte la forma.

$$S_{ia} - S_{ib} = b P_i + c_a X_{ia} + c_b X_{ib} + \mu_{ia} - \mu_{ib}$$

Incluso si algunas (o todas) las variables de X no varían en el tiempo, aún se puede permitir que X afecte los cambios en el tiempo en la escolaridad.

El método de correspondencia de puntuación de la propensión antes analizado puede ayudar a asegurar que el grupo de comparación sea similar al grupo de tratamiento antes de realizar la doble diferencia. En un interesante estudio de un programa de empleos estadounidense, se concluyó que el no asegurar que las comparaciones se hubieran efectuado en un área de apoyo común fue una fuente importante de sesgo en la estimación de doble diferencia en comparación con un grupo de control aleatorizado. Sin embargo, dentro del área de apoyo común, el sesgo dependiente de X no varió mucho en el tiempo. Por lo tanto, es razonable tomar la doble diferencia después de realizar la correspondencia (véase Heckman y otros (1998)).

En la práctica, sin embargo, puede ser difícil hacer un seguimiento de los hogares de las encuestas. Podría no ser fácil encontrar a todos los hogares que se incluyeron originalmente en la encuesta básica. Algunas personas de la encuesta básica podrían no desear ser entrevistadas de nuevo o podrían haberse trasladado a una ubicación desconocida.

Si los retiros de la muestra son exclusivamente aleatorios, la encuesta de seguimiento todavía será representativa de la misma población de la encuesta básica. Sin embargo, si existe alguna tendencia sistemática de las personas con ciertas características a abandonar la muestra, esto sí representaría un problema. Esto se llama “sesgo de desgaste o de disminución normal”. Por ejemplo, PROSCOL podría ayudar a algunas familias pobres a trasladarse a una vivienda mejor. Además, incluso cuando la selección de los participantes sólo se base en información disponible en la fecha básica de referencia o alrededor de ella (el efecto 0_i constante en el tiempo), los participantes seleccionados bien podrían retirarse en forma voluntaria sobre la base de cambios después de esa fecha. Ese desgaste del grupo de tratamiento claramente sesgará una estimación de doble diferencia del impacto del programa. En el Recuadro 3.3 se resumen los pasos para realizar una estimación de doble diferencia.

Recuadro 3.3. Ejecutando una doble diferencia

El método de doble diferencia supone comparar un grupo de tratamiento con un grupo de comparación (que idealmente se determinaría por medio del método de correspondencia del Recuadro 3.2) antes y después de la intervención. Los pasos principales son los siguientes:

Paso 1: Se necesita una encuesta básica antes de establecer la intervención y la encuesta debe cubrir tanto a participantes como a no participantes. Si no se sabe quién participará, se debe realizar una estimación bien informada. Se deberá conversar con los administradores del programa.

Paso 2: Luego se requerirán una o más encuestas de seguimiento después del establecimiento del programa. Éstas deben ser muy comparables a las encuestas básicas (en términos del cuestionario, la entrevista, etc.). Idealmente, las encuestas de seguimiento deben incluir las mismas observaciones muestreadas que la encuesta básica. Si no es posible, deben corresponder a los mismos grupos geográficos o estratos con respecto a alguna otra variable.

Paso 3: Hay que calcular la diferencia media entre los valores después y antes del indicador de resultados para cada uno de los grupos de tratamiento y comparación.

Paso 4: Se debe calcular la diferencia entre estas dos diferencias medias. Esa es la estimación del impacto del programa.

Esta es la versión más simple de la doble diferencia. También se podría necesitar controlar por las diferencias en las condiciones iniciales exógenas o los cambios en las variables exógenas, posiblemente dando cabida a los efectos de interacción con el programa (de modo que la ganancia producida por la intervención del programa sea alguna función de las variables observables). Un modelo de regresión adecuado puede admitir estas variaciones.

Uso de variables instrumentales

Dado que no hay encuesta básica de los mismos hogares para emplear el método de doble diferencia, el profesor recomienda otra metodología para obtener una estimación que sea sólida para los elementos no observables problemáticos: una “variable instrumental”.

Una variable instrumental es la clásica solución para el problema de un regresor endógeno. Una variable instrumental es una fuente observable de variación exógena de la participación en el programa. En otras palabras, está correlacionada con P , pero aún no está en la

regresión para la escolaridad y no está correlacionada con el término de error en la ecuación de escolaridad, ε . De modo que se debe tener al menos una variable en Z que no esté en X y que no esté correlacionada con ε . Luego, la estimación de variables instrumentales del impacto del programa se obtiene reemplazando P por su valor predicho dependiente de Z . Puesto que este valor predicho depende sólo de Z (que es exógeno) y Z está correlacionado con ε , ahora es razonable aplicar los MCO a esta nueva regresión.

Puesto que los valores predichos sólo dependen de la variación exógena debido a la variable instrumental y a las demás variables exógenas, los no observables ya no son problemáticos porque estarán correlacionados con el término de error en la regresión de escolaridad. Esto también sugiere otra forma más eficiente de tratar el problema. Cabe recordar que la fuente de sesgo en la estimación anterior del impacto del programa era la correlación entre el término de error de la ecuación de escolaridad y el correspondiente a la ecuación de participación. Esto es lo que crea la correlación entre la participación y el término de error en la ecuación de la escolaridad. Así, una forma natural de deshacerse del problema cuando se tiene una variable instrumental es agregar los valores residuales de la ecuación de participación de la primera etapa a la ecuación de escolaridad, pero manteniendo la participación real en la regresión de escolaridad. Sin embargo, puesto que ahora se ha agregado el valor estimado del término de error de la ecuación de participación a la regresión de escolaridad, es posible tratar la participación como exógena y ejecutar los MCO. Esto sólo funciona si hay un instrumento válido. Si no lo hay, la regresión no realizará la estimación, porque en un modelo lineal el valor residual de la participación será perfectamente predecible a partir de la participación real y X .

Una variable instrumental también puede ayudar si hay un error de medición apreciable en los datos de participación en el programa, otra posible fuente de sesgo. Un error de medición significa que existe la posibilidad de que la participación en el programa varíe más de lo que realmente hace. Naturalmente, esta sobrestimación en la variación de P produce una subestimación de su coeficiente b . Esto se denomina sesgo de atenuación, porque este sesgo atenúa el coeficiente de regresión estimado.

Aunque una variable instrumental puede ser extremadamente útil, en la práctica es necesario ser prudente. Cuando la participación real sólo se sustituye por su valor predicho y se ejecutan los MCO, esto no entregará los errores estándar correctos: la computadora no sabrá que se deben usar parámetros previamente estimados para obtener los valores predichos. Se requiere una corrección a los errores estándar de MCO, aunque existen paquetes estadísticos que permiten hacerlo fácilmente, al menos para los modelos lineales.

Sin embargo, si hubiera una variable dependiente que sólo pudiera adoptar dos posibles valores, por ejemplo, en la escuela o no en la escuela, se debería usar un modelo de respuesta binario no lineal, como logit o probit. El principio para probar la exogeneidad de la participación en el programa es similar en este caso. Existe un estudio por Rivers y Vuong (1988) en que se analiza el problema para estos modelos; en Blundell y Smith (1993) se proporciona una visión general útil de diversos modelos no lineales en los cuales existe un regresor endógeno.

Pruebas de las metodologías

Cuando el analista comienza a pensar sobre la identificación de una variable instrumental, comprende que no es un proceso directo. Todas las posibilidades en que ha pensado también se podrían incluir junto a las variables en X . El problema es encontrar una “restricción de exclusión” válida que justifique colocar alguna variable en la ecuación de participación, pero no en la ecuación de escolaridad.

El analista decide intentar el método de correspondencia de puntuación de la propensión. El modelo logit de participación parece bastante sensible y sugiere que PROSCOL está bien focalizado. Casi todas las variables de las que se podría esperar una asociación con la pobreza tienen coeficientes positivos y significativos. Luego el analista realiza la correspondencia de puntuación de la propensión. En una comparación de las tasas de matrícula escolar medias, los resultados muestran que los niños del grupo de comparación pareado tenían una tasa de matrícula del 60%, en comparación con el 80% para las familias del PROSCOL.

Para responder al problema del ingreso no percibido, el análisis recurre a una encuesta existente del trabajo infantil en que se consultaba sobre los ingresos (en este país en desarrollo, existe una prohibición oficial de que los niños menores de 16 años trabajen, pero al gobierno le ha resultado difícil hacerla cumplir; no obstante, los salarios infantiles son un problema delicado). Según esta encuesta, se pueden determinar los ingresos que habría recibido un niño si no hubiera asistido a la escuela.

Luego es posible restar el monto del ingreso no percibido del pago en efectivo de PROSCOL a los participantes y calcular así la transferencia neta de ingresos. Al restar esta transferencia neta del ingreso total, será posible calcular de dónde provienen los participantes en PROSCOL en la distribución de ingresos previa a la intervención. No son tan pobres como se pensaba al principio (si se ignora el ingreso no percibido), pero siguen siendo pobres; por ejemplo, dos tercios de ellos están bajo la línea de pobreza oficial del país.

Habiendo calculado la ganancia de ingreso neta de todos los participantes, ahora es posible calcular la tasa de pobreza con y sin PROSCOL. En términos simples, la tasa de pobreza posterior a la intervención (con el programa) es la proporción de la población que vive en hogares con un ingreso por persona bajo la línea de pobreza, donde “ingreso” es el ingreso observado (incluidos los ingresos de transferencia bruta de PROSCOL). Esto se puede calcular directamente a partir de la encuesta domiciliaria. Al sustraer la ganancia de ingreso neta (la transferencia de efectivo de PROSCOL menos los ingresos no percibidos del trabajo infantil) atribuido a PROSCOL de todos los ingresos observados, los resultados muestran una nueva distribución de los ingresos anteriores a la intervención. Entonces, la tasa de pobreza sin el programa es la proporción de personas que viven en hogares pobres, según esta nueva distribución. El análisis concluye que la tasa de pobreza observada del 32% en el noroeste habría sido un 36% si no hubiera existido PROSCOL. El programa permite que ahora un 4% de la población se libere de la pobreza. Las ganancias en escolaridad significan que también habrá ventajas pecuniarias y no pecuniarias para los

pobres en el futuro. En el proceso de medición de la pobreza, el analista recuerda haber aprendido que la proporción de personas bajo la línea de pobreza sólo es una medida básica, porque no indica nada sobre los cambios bajo la línea (véase el Recuadro 3.4). Luego calcula el índice de la brecha de pobreza y el índice de la brecha de pobreza al cuadrado y los resultados sugieren que éstos también han descendido como resultado del PROSCOL.

Recuadro 3.4 Mediciones de pobreza

La medida más simple y más común de la pobreza es el índice de recuento de la pobreza. En este caso, es la proporción de la población que vive en hogares con un ingreso por persona bajo la línea de pobreza (en otros países es una medida basada en el consumo, lo que tiene ciertas ventajas; para un análisis y referencias, véase Ravallion, 1994).

El índice de recuento de la pobreza no indica nada sobre la distribución del ingreso bajo la línea de pobreza: una persona pobre puede quedar en peor situación, pero el índice de recuento de la pobreza no variará ni reflejará los progresos entre los pobres, a menos que crucen la línea de pobreza.

Una alternativa para el índice de recuento de la pobreza, utilizada en forma generalizada, es el índice de brecha de pobreza (PG). La brecha de pobreza para cada hogar es la diferencia entre la línea de pobreza y el ingreso del hogar; para aquellos sobre la línea de pobreza, la brecha es cero. Cuando la brecha de pobreza se normaliza por línea de pobreza y se calcula su media en todos los hogares (pobres o no), se obtiene el índice de brecha de pobreza.

El índice de brecha de pobreza indicará el nivel en que el programa ha influido en la profundidad de la pobreza, pero no reflejará ningún cambio en la distribución entre los pobres causado por el programa. Por ejemplo, si el programa supone un pequeño avance para una persona pobre que se encuentra sobre el ingreso medio de los pobres, a expensas de una pérdida equivalente de alguien bajo esa media, el índice de brecha de pobreza no variará.

Existen diversas mediciones “sensibles a la distribución” que reflejarán esos cambios en la distribución entre los pobres. Una de estas mediciones es la “brecha de pobreza al cuadrado” (Foster, Greer y Thorbecke, 1984). Se calcula de la misma forma que la brecha de pobreza, excepto que cada brecha de pobreza como proporción de la línea de pobreza se eleva al cuadrado antes de sacar la media (nuevamente entre pobres y no pobres). Otro ejemplo de una medición de pobreza sensible a la distribución es el índice Watts. Es la media del logaritmo de la razón entre la línea de pobreza y el ingreso, donde la razón se fija en uno para los no pobres. Atkinson (1987) describe otros ejemplos en sus estudios.

En este cálculo, el analista también reconoce que hay cierta incertidumbre acerca de la línea de pobreza del país. Para probar los resultados, repite el cálculo sobre una amplia variedad de líneas de pobreza y concluye que en una línea de pobreza para la cual el 50% de la población es pobre (según los ingresos observados después de la intervención), la proporción habría sido del 52% sin PROSCOL. En una línea de pobreza en que el 15% no es cubierto por el programa, la proporción habrá sido del 19% sin éste. Mediante la repetición de estos cálculos en toda la gama de ingresos, se trazaron todas las “curvas de incidencia de la pobreza”, con el programa y sin él. Esto también se denomina “función de distribución acumulativa” (véase el Recuadro 3.5).

Recuadro 3.5 Comparación de la pobreza con el programa y sin él

Con la ayuda de los métodos descritos en el texto principal y los recuadros anteriores, se obtiene una estimación de los beneficios para cada hogar. En las evaluaciones más simples es sólo un número. Pero es mejor dejar que varíe con las características del hogar. Luego se puede resumir esta información en forma de curvas de incidencia de la pobreza (PIC) con el programa y si él.

Paso 1: Ya debería existir el ingreso posterior a la intervención (u otro indicador de bienestar) para cada hogar de toda la muestra (que comprenda participantes y no participantes); éstos son datos. También se sabe el número de personas que vive en cada hogar. Y, por cierto, se conoce el número total de personas de la muestra (N; o podría ser el tamaño estimado de la población, si se han usado tasas de muestreo inversas para “agotar” la observación de cada muestra).

Paso 2: Se puede representar esta información en forma de una curva PIC. Esto proporciona (en el eje vertical) el porcentaje de la población que vive en hogares con un ingreso menor o igual al valor del eje horizontal. Para trazar este gráfico, se puede comenzar con el hogar más pobre, marcar su ingreso en el eje horizontal y luego contar en el eje vertical 100 veces el número de personas de ese hogar dividido por N. El siguiente punto es la proporción que vive en los dos hogares más pobres, y así sucesivamente. Con esto se obtiene la PIC posterior a la intervención.

Paso 3: Ahora hay que calcular la distribución de ingreso previo a la intervención. Para obtenerla, se resta la ganancia estimada para cada hogar de su ingreso posterior a la intervención. Luego hay una lista de ingresos posteriores a la intervención, uno para cada hogar de la muestra. Repetir entonces el Paso 2. Se obtiene la PIC previa a la intervención.

Si se considera cualquier nivel de ingreso dado del eje horizontal como una línea de pobreza, la diferencia entre las dos PIC en ese punto da el impacto en el índice de recuento de la pobreza para esa línea de pobreza (Recuadro 3.4). Alternativamente, al examinar el eje horizontal se obtiene el aumento de ingreso en ese percentil. Si ninguna de las ganancias es negativa, la PIC posterior a la intervención debe quedar bajo la respectiva curva antes de la intervención. La pobreza habrá disminuido, sin importar qué línea de pobreza se esté utilizando. Por cierto, esto también se mantiene para una amplia clase de medidas de pobreza; véase Atkinson (1987). Si algunas ganancias son negativas, las PIC se intersectarán. En ese caso, la comparación de la pobreza es ambigua; la respuesta dependerá de las líneas de pobreza y medidas de pobreza que se usen (un análisis más detallado se encuentra en Ravallion, 1994). Se podrían aplicar restricciones a priori a la gama de líneas de pobreza admisibles. Por ejemplo, se podría confiar en que la línea de pobreza no supere cierto valor máximo y si se produce la intersección sobre ese valor, la

comparación de la pobreza sería inequívoca. Si el punto de intersección (y podría haber más de uno) está bajo la línea de pobreza máxima admisible, sólo sería posible una comparación de pobreza sólida para un conjunto restringido de medidas de pobreza. Para verificar el grado de restricción que debe tener el conjunto, se pueden calcular las curvas de profundidad de la pobreza (PDC). Es posible obtenerlas formando simplemente la suma acumulativa hasta cada punto de la PIC (de modo que el segundo punto de la PDC sea el primer punto de la PIC más el segundo punto, y así sucesivamente).

Si las PDC no se intersectan, el impacto del programa en la pobreza es inequívoco, siempre y cuando la atención quede restringida al índice de brecha de pobreza o a cualquiera de las medidas de pobreza sensibles a la distribución descritas en el recuadro 3.4. Si las PDC se intersectan, se pueden calcular las “curvas de severidad de la pobreza” con el programa y sin él, formando las sumas acumulativas bajo las PDC. Si estas no se intersectan sobre la gama de líneas de pobreza admisibles, el impacto en cualquiera de las medidas de pobreza sensibles a la distribución del Recuadro 3.4 es inequívoco.

Incorporación de información del terreno

En la implementación de cada programa, existen conocimientos de los beneficiarios y administradores del programa que podrían estar o no reflejados en los datos del programa. Por ejemplo, en este caso la percepción de los que trabajan en terreno es que la mayoría de las familias de PROSCOL son pobres y que el programa desde luego sirve de ayuda. Cuando el analista trata esto con un sociólogo que trabaja en el programa, percibe cierta incertidumbre en la realidad del ingreso no percibido y el tema del trabajo. El sociólogo comenta que en el terreno se observan muchos niños de familias pobres que trabajan y asisten a la escuela y que algunos de los niños más pequeños que no asisten a la escuela

parecen no estar trabajando. El analista comprende que esto requiere la verificación de si existe alguna diferencia en la cantidad de trabajo realizada por los niños dentro de PROSCOL con respecto al realizado por un grupo de comparación pareado. Sin embargo, estos datos no están disponibles en la encuesta domiciliaria, aunque sería posible presentar los resultados con la deducción del ingreso no percibido y sin ella.

El sociólogo también ha observado que para que una familia pobre ingrese a PROSCOL, tiene mucha importancia el área de la junta escolar (SBA, *school-board area*) donde ésta viva. Todas las SBA obtienen una asignación en PROSCOL desde el centro, incluso aquellas que tienen muy pocas familias pobres. Si se es pobre, pero se habita dentro de una SBA próspera, es más probable obtener ayuda de PROSCOL que si se vive en una SBA pobre. Por consiguiente, lo que realmente importa es la pobreza relativa —relativa a las demás familias del área en que se vive—, la que influye mucho más que el nivel de vida absoluto.

Esta asignación influiría en la participación en PROSCOL, pero no es dable esperar que tuviera importancia para la asistencia a la escuela. Esta última dependería más del nivel de vida absoluto, las circunstancias familiares y las características de la escuela. Así, la asignación presupuestaria de PROSCOL en las SBA se puede usar como variable instrumental para eliminar el sesgo en las estimaciones del impacto del programa.

Hay información acerca de la SBA a la que pertenece cada hogar en la encuesta domiciliaria, las normas aplicadas por el centro en la asignación de fondos de PROSCOL a las SBA y el monto que el centro ha asignado a cada SBA. Las asignaciones se basan en el número de niños en edad escolar, con un “factor de ajuste” correspondiente al grado de pobreza estimado para la SBA. Sin embargo, la norma es algo vaga.

El analista intenta tomar en cuenta estos puntos y vuelve a ejecutar la regresión para la escolaridad, pero reemplaza la participación real en PROSCOL por su valor predicho (la puntuación de la propensión) de la regresión para la participación, que ahora incluye la asignación presupuestaria a la SBA. Es útil tener de antemano tantas características como sea posible de las escuelas en la regresión correspondiente a la asistencia. Aunque las características de la escuela oficialmente parecen no influir en la forma en que se asignan los recursos de PROSCOL, cualquier característica escolar omitida que influya en forma simultánea en las asignaciones de PROSCOL por SBA y en cada resultado de escolaridad dejará un sesgo en las estimaciones de variables instrumentales del analista. Aunque siempre existe la posibilidad de sesgo, con abundantes variables de control geográfica este método debería ofrecer al menos un comparador creíble para la estimación de correspondencia.

De los resultados se desprende que la asignación de presupuesto a la SBA efectivamente tiene un coeficiente positivo significativo en la regresión logit para la participación en PROSCOL. Ahora la participación en PROSCOL (predicha) es significativa en una regresión para la matrícula escolar, en la cual se incluyen las mismas variables de la regresión logit, excepto la asignación de presupuesto de la SBA. El coeficiente implica que la tasa de matrícula es 15 puntos porcentuales más alta para los participantes en PROSCOL

de lo que habría sido de otro modo. También se ejecuta una regresión para los años de escolaridad, para niños y niñas por separado. Para niños y niñas de 18 años, los resultados indican que habrían abandonado la escuela casi dos años antes si no hubiera sido por PROSCOL. Sin embargo, esta regresión plantea preguntas sobre si se están utilizando los errores estándares correctos y si se deben usar modelos lineales.

Planificación del trabajo futuro

Por último, el analista está preparado para presentar un informe con los resultados de las evaluaciones. Estos resultados indican que PROSCOL está funcionando bastante bien y debido a ello, las autoridades responsables se muestran interesadas en ampliar el programa. El analista tiene algunas observaciones importantes que hacer, las que han surgido del proceso por el que ha pasado para llevar a cabo la evaluación:

- La evaluación del impacto puede ser mucho más difícil de lo previsto al principio;
- Es posible obtener una gama de estimaciones, preocupante por su extensión, según las especificaciones de la metodología usada;
- Es apropiado usar métodos alternativos en las situaciones, por cierto frecuentes, de contar con datos poco adecuados, aunque cada método tiene dificultades, y
- Se debe ser ecléctico con respecto a los datos.

Además de las lecciones que ha aprendido el analista, tiene algunas recomendaciones clave para el futuro trabajo de evaluación de PROSCOL. En primer lugar, sería conveniente excluir aleatoriamente algunas familias calificadas dentro de PROSCOL en el resto del país y luego realizar una encuesta de seguimiento de los participantes reales y los excluidos en forma aleatoria de la participación. Esto entregaría una estimación más precisa de los beneficios. Sin embargo, sería políticamente delicado excluir a algunos. No obstante, si el programa no cuenta con suficientes recursos para cubrir todo el país en una sola etapa y se deberá optar por quienes serán seleccionados primero, por cierto sería preferible efectuar esa selección en forma aleatoria entre los participantes calificados. Alternativamente, sería posible seleccionar las escuelas o las áreas de juntas escolares en forma aleatoria en la primera etapa. Sin duda, esto haría de la elección de las escuelas o áreas de junta escolar una buena variable instrumental para la asignación de cada programa.

En segundo lugar, si esto no es factible, se recomienda realizar una encuesta básica en las áreas en las cuales es probable que haya una elevada concentración de participantes en PROSCOL antes que comience el programa en el sur. Esto se podría realizar simultáneamente con la siguiente etapa de la encuesta nacional que se utilizó para evaluar el programa PROSCOL. También sería conveniente agregar algunas preguntas a la encuesta, por ejemplo, si los niños realizan algún trabajo remunerado.

Y en tercer lugar, sería útil incluir trabajo cualitativo, para ayudar a plantear las hipótesis

que se deben probar y evaluar la credibilidad de los supuestos clave establecidos en el análisis cuantitativo.

Nota

1. En Heckman, Lalonde y Smith (1999) y Abadie, Angrist e Imbens (1998) se encuentra un análisis sobre los efectos del tratamiento en los cuartiles.

Capítulo 4

Cómo aprovechar las evaluaciones de impacto de “Práctica Óptima”*

En los capítulos anteriores se presentaron los métodos, problemas y dificultades clave que pueden surgir al evaluar el impacto de un proyecto. Al revisar los estudios de casos indicados en el Cuadro 4.1, surgen muchos ejemplos ilustrativos de interesantes enfoques frente al diseño, uso de datos, elección y aplicación de los métodos analíticos usados y la creación de capacidades en el país. A continuación se destacan estos ejemplos, además de un análisis de los costos de las evaluaciones y los problemas de economía política que podrían surgir en la implementación.

Los 15 estudios de casos incluidos en la revisión se seleccionaron de una variedad de evaluaciones realizadas por el Banco Mundial, otras agencias de donantes, instituciones de investigación y empresas consultoras privadas. Se seleccionaron como ejemplos de una “práctica óptima” por su rigor metodológico, intentando reflejar una gama de ejemplos de diferentes sectores y regiones. Aunque cada evaluación de impacto tiene sus fortalezas y debilidades, las lecciones aprendidas de estas experiencias deberían servir de ayuda a los administradores de proyectos o a analistas políticos que intentan diseñar e implementar un futuro trabajo.

Planificación anticipada y cuidadosa del diseño de la evaluación

Una preparación cuidadosa durante las etapas iniciales de la identificación del proyecto asegurará que se recopile la información correcta y que se puedan emplear los resultados para hacer ajustes a los componentes del proyecto a mitad del proceso. Mediante una planificación anticipada y cuidadosa, es posible incorporar todos los elementos que contribuirán a una evaluación de impacto rigurosa, como una encuesta básica (de referencia) con un grupo de control aleatorio y datos cualitativos sobre los procesos que podrían influir en los efectos.

Proyecto de Nutrición y Desarrollo Preescolar de Uganda. Esta evaluación, aunque aún no se implementa, representa un excelente ejemplo de una planificación anticipada y cuidadosa (véase Banco Mundial, 1998a; García, Alderman y Rudqvist, 1999). El proyecto mismo se centra en el fortalecimiento de la capacidad de los padres y comunidades para cuidar a los niños, proporcionándoles conocimientos sobre mejores prácticas de cuidado infantil y mejorando las oportunidades de aumentar el ingreso. Se basa en la comunidad y lo aplica una red de organizaciones no gubernamentales (ONG). El componente de evaluación, que se integró al ciclo del proyecto desde el primer día, se acerca al ideal en

* Este capítulo se basa en gran medida en los estudios de casos de práctica óptima del Anexo I y en publicaciones de información general preparados por Gillette Hall y Julia Lane, y Subbarao y otros (1999).

términos de su diseño. En primer lugar, se generan datos de encuestas básicas y de seguimiento, junto con un grupo de control aleatorio, de modo que se puede evaluar en forma rigurosa el efecto del programa a nivel de los beneficiarios. En segundo lugar, mejora el componente cuantitativo con un proceso participativo (cualitativo) de supervisión y evaluación.

En el aspecto cuantitativo, el proyecto se diseñó con miras a un diseño de estudio experimental, en que las subdivisiones se asignarán en forma aleatoria a los grupos de tratamiento y control. Luego se usarán tarjetas de salud para registrar la información sobre el peso de los niños en las subdivisiones de tratamiento y control. Además, la encuesta básica domiciliaria se realizará antes de entregar los servicios a las comunidades, al igual que una encuesta de seguimiento de los mismos hogares dos años después. Se espera que una rápida revisión de estos datos sirva de guía para la decisión de ampliar algunos componentes de la intervención durante la revisión a mitad del proyecto. Un análisis más profundo de los datos al término del proyecto orientará el diseño de la segunda fase del proyecto.

Proyecto de crédito con educación de Ghana. La evaluación de este proyecto fue muy compleja y tuvo varios pasos intermedios. En el proyecto se combinan los elementos de un plan de préstamo grupal con la educación sobre nociones básicas en salud, nutrición, planificación y espaciamiento de los nacimientos y capacitación en pequeños negocios. En general, la evaluación se centra en analizar el estado nutricional de los niños, la capacidad económica de las mujeres para invertir en alimentos y atención de salud, conocimientos y adopción de la lactancia materna y destete por parte de las mujeres. Comienza con un marco conceptual muy claro, el que se ilustra más adelante. Este esquema detalla claramente los aportes, beneficios intermedios y resultados a largo plazo, en una forma que facilita la elaboración de diversos modelos y su interpretación. Al planificar en forma cuidadosa la evaluación y trabajar con un esquema desde una etapa inicial, fue posible aclarar muchos aspectos de un diseño relativamente complejo (véase el Anexo 1.6).

Enfoques para la evaluación cuando no hay datos básicos

En la práctica, muchas evaluaciones no cuentan con los datos adecuados. Se agregan evaluaciones después que es posible realizar una encuesta básica o en ausencia de grupos de comparación. Algunos ejemplos de este aspecto son las evaluaciones de los proyectos Alimentos por Educación de Bangladesh, PROBECAT de México, Programas de Trabajo Activo de la República Checa y TRABAJAR de Argentina. Sin datos básicos, se deben crear controles usando los métodos de pareo analizados en los capítulos anteriores. Sin embargo, esto puede ser bastante complicado. Un buen ejemplo lo constituye la técnica de correspondencia de puntuación de la propensión usada en el proyecto TRABAJAR de Argentina para crear un grupo de control con datos transversales de los participantes y no participantes en el programa.

Cuadro 4.1 Resumen de evaluaciones de impacto de “práctica óptima”

<i>Proyecto:</i>	<i>País</i>	<i>Tipo de base de datos</i>	<i>Unidad de análisis</i>	<i>Medidas de resultados</i>	<i>Método econométrico</i>				<i>Cualitativo</i>	<i>Fortalezas</i>	
					<i>Aleatorización</i>	<i>Pareo</i>	<i>Comparaciones reflexivas</i>	<i>Doble diferencia</i>			<i>VARIABLES instrumentales</i>
Educación											
Radio Nicaragua	Nicaragua	Encuesta básica y posterior a la intervención	Estudiantes y clases	Calificaciones en las pruebas	Sí	No	Sí	No	No	No	Diseño de cuestionario
Reforma de autonomía escolar	Nicaragua	Evaluaciones cualitativas y panel de encuestas	Estudiantes, padres, maestros, directores	Calificaciones en las pruebas, grado de toma de decisiones local	No	Sí	Sí	No	No	Sí	Combinación cualitativa y cuantitativa
Libros de texto	Kenya	Encuesta básica y posterior a la intervención	Estudiantes, clases, maestros	Calificaciones en las pruebas	Sí	No	Sí	Sí	Sí	No	Análisis de factores distractivos
Abandono	Filipinas	Encuesta básica y posterior a la intervención	Estudiantes, clases, maestros	Calificaciones en las pruebas y nivel de deserción	Sí	No	Sí	Sí	Sí	No	Análisis de costos/beneficios; creación de capacidades
Programas laborales											
TRABAJAR	Argentina	Encuesta domiciliaria, censo, registros administrativos, evaluaciones sociales	Trabajadores, hogares	Ingreso, focalización, costos	No	Sí	No	No	Sí	Sí	Uso razonable de las fuentes de datos existentes, técnicas analíticas innovadoras
PROBECAT	México	Encuestas retrospectivas y de la fuerza laboral	Trabajadores	Resultados en el ingreso y el empleo	No	Sí	No	No	No	No	Técnica de correspondencia

Programas de Trabajo Activo	República Checa	Encuestas retrospectivas por correo	Trabajadores	Resultados en el ingreso y el empleo	No	Sí	No	No	No	No	Técnica de correspondencia
Finanzas											
Microfinanzas	Bangladesh	Encuesta posterior a la intervención y registros administrativos	Hogares	Consumo y educación	Sí	Sí	No	Sí	No	No	Análisis de factores distractivos
Crédito con educación	Ghana	Encuesta básica y posterior a la intervención	Pares madre e hijo	Ingreso, salud y habilitación	Sí	Sí	Sí	No	No	Sí	Uso de información cualitativa y cuantitativa
Financiamiento de la salud	Níger	Encuesta básica y posterior a la intervención, más registros administrativos	Hogares y centros de salud	Recuperación de costos y acceso	No	Sí (en distritos)	Sí	No	No	No	Uso de datos administrativos
Alimentos y nutrición											
Alimentos para la educación	Bangladesh	Encuesta sobre gastos familiares	Hogares y comunidades	Asistencia escolar	No	No	No	No	Sí	No	Uso imaginativo de instrumentos para abordar el problema de selección con datos estándar
Salud, educación y nutrición	México	Encuestas básicas y posteriores a la intervención	Hogares	Resultados en salud, educación y nutrición	Sí	Sí	Sí	Se ignora	Se ignora	No	Conceptualización clara; análisis de factores distractivos

Infraestructura											
Fondo de Inversión Social	Bolivia	Encuestas básicas y de seguimiento	Hogares, proyectos	Indicadores de educación y salud	Sí	Sí	Sí	Sí	Sí	No	Aplicación de una variedad de metodologías de evaluación
Caminos rurales	Vietnam	Encuestas básicas y de seguimiento	Hogares, comunidades	Indicadores de bienestar a nivel de hogar y comunidad	No	Sí	Sí	Sí	Sí	No	Mide resultados en el bienestar
Agricultura											
Proyecto de extensión nacional	Kenya	Evaluaciones de beneficiarios, datos de panel	Hogares, granjas	Productividad y eficiencia de granjas	No	No	Sí	No	No	No	Aplicabilidad de los resultados en la política

El proyecto TRABAJAR II en Argentina. Este proyecto se centró en proporcionar empleos de bajo salario en pequeños subproyectos de infraestructura social y económica seleccionados por agrupaciones locales. La evaluación de impacto del programa se diseñó para evaluar si los ingresos de los participantes en el programa eran más altos de lo que hubieran sido en ausencia del programa. Los métodos que se usan habitualmente para estimar el ingreso de los hogares sin la intervención no eran factibles de aplicar en el caso del programa TRABAJAR: no se había realizado aleatorización para crear un grupo de control que se pudiera usar para comparar el ingreso de los beneficiarios del proyecto y no se contaba con una encuesta básica, lo que eliminaba la posibilidad de realizar una evaluación antes y después.

En su lugar, en la evaluación de TRABAJAR se utilizaron los datos existentes para crear un grupo de control a través de la correspondencia entre participantes del programa y no participantes de la población nacional para un conjunto de variables socioeconómicas, como escolaridad, género, vivienda, percepciones subjetivas del bienestar y afiliación a partidos políticos y asociaciones de vecinos, usando una técnica llamada puntuación de la propensión. El estudio es un ejemplo del uso imaginativo de los datos existentes de una encuesta domiciliaria a nivel nacional —la Encuesta de Desarrollo Social (EDS)— para generar el grupo de comparación, junto con una encuesta más reducida de participantes en TRABAJAR realizada específicamente para fines de la evaluación. La encuesta más reducida se diseñó cuidadosamente, de modo de utilizar el mismo cuestionario que la EDS y los mismos equipos de entrevistadores y se llevó a cabo aproximadamente al mismo tiempo, para poder realizar el ejercicio de correspondencia o pareo. Esta técnica fue posible en el caso de TRABAJAR, porque se estaba realizando una encuesta domiciliaria a nivel nacional y los evaluadores podían aprovechar esta encuesta para sobremuestrear los participantes en TRABAJAR. Se usaron los mismos equipos de entrevistadores para la encuesta nacional y la encuesta del proyecto, lo que arrojó ganancias en eficiencia en la recopilación de datos (véase el Anexo 1.1).

Evaluación de programas del mercado laboral en la República Checa. En esta evaluación se intentó cubrir cinco programas de trabajo activo para (a) determinar si los participantes en los distintos programas tenían más éxito para reincorporarse al mercado laboral que los que no participaban en el programa y si esto variaba a través de los subgrupos y con las condiciones del mercado laboral y (b) determinar la eficacia en función de los costos de cada programa y sugerir mejoramientos. En esta evaluación se utilizó una técnica de correspondencia, porque no se recopilaron datos básicos (de referencia). Los evaluadores entrevistaron a los participantes y luego seleccionaron una muestra aleatoria de no participantes. Puesto que los no participantes eran sistemáticamente de mayor edad y con un menor nivel de instrucción, los evaluadores tuvieron que crear un grupo de comparación razonable para cada programa. Para ello, se tomó a cada participante de a uno y se le comparó con cada individuo del grupo de no participante sobre la base de siete características: edad, género, educación, número de meses empleado antes del registro, tamaño del pueblo, estado civil y tipo del último empleo. La correspondencia más cercana se introdujo en el grupo de comparación. Aunque este enfoque es directo, existe la

posibilidad de sesgo en la selección, esto es, que el grupo de no participantes sea sistemáticamente diferente al grupo de participantes sobre la base de elementos no observables (Anexo 1.5).

El manejo de las restricciones al elaborar controles adecuados

A veces es posible hacer una selección aleatoria o contar con controles experimentales, pero no es factible en términos de política. En este caso, se puede efectuar la aleatorización aprovechando cualquier plan para probar experimentalmente el proyecto en ciertas áreas restringidas. Inicialmente se pueden seleccionar aleatoriamente las áreas en las cuales se aplicará el proyecto piloto, fijando posibles áreas futuras del proyecto como controles. En el tiempo se pueden incluir aleatoriamente comunidades adicionales en el proyecto. Tres ejemplos ilustran la forma en que se puede manejar una situación en la cual la aleatorización no era factible en términos de política o debido a otros aspectos. En Vietnam se evaluará un proyecto de transporte rural con información limitada y sin aleatorización. El Fondo Hondureño de Inversión Social es un buen ejemplo de la creación de un grupo de control en proyectos guiados por la demanda, usando una comparación pareada ex post basada en un corte transversal único de datos. Evaluar proyectos guiados por la demanda puede ser especialmente difícil, dado que no se conocen anticipadamente los proyectos o comunidades que participarán en el proyecto. Y tercero, la evaluación del Fondo de Inversión Social de Bolivia en la Región del Chaco proporciona un buen ejemplo de cómo incorporar la aleatorización en proyectos guiados por la demanda de forma que permita su orientación hacia grupos destinatarios específicos.

El Proyecto de caminos rurales de Vietnam. Este proyecto está destinado a reducir la pobreza en las áreas rurales mejorando el acceso a las comunidades rurales y conectándolas con las redes viales del distrito y la provincia. El diseño de la evaluación del impacto se centra en los datos de encuestas básicas y de seguimiento recopilados para una muestra de comunidades del proyecto y del grupo de comparación identificado mediante técnicas de comparación pareada. De una muestra aleatoria de miembros de la comunidad dentro del proyecto (tratamiento) y fuera de éste (comparación), se recopilará información básica y posterior a la intervención sobre indicadores como rendimientos de la producción agrícola a nivel de la comunidad, diversificación de la fuente de ingresos, oportunidades de empleo, disponibilidad de bienes, servicios e instalaciones y patrimonio y distribución de activos. Estos datos se emplearán para comparar los cambios en los resultados antes y después de la intervención entre comunidades del proyecto y fuera de éste, usando la técnica de “doble diferencia”.

Idealmente, las comunidades de tratamiento y de comparación deberían ser equivalentes en todas sus características observadas y no observadas, siendo la única diferencia entre ellas que las comunidades de tratamiento se benefician con la intervención, mientras que las comunidades de comparación no. Puesto que no se había realizado una asignación aleatoria a grupos de tratamiento y de comparación y no se disponía de los datos necesarios para tomar decisiones fundamentadas sobre los controles adecuados en el momento de la selección de la muestra, se tomaron muestras aleatorias de comunidades dentro del proyecto

y fuera de éste. Específicamente, se seleccionaron las comunidades del proyecto de una lista de todas las comunidades con proyectos propuestos en cada provincia. Luego, se seleccionaron comunidades de comparación de una lista de todas las comunidades restantes sin proyectos propuestos, pero en los mismos distritos que las comunidades de tratamiento. Con la información recopilada para la evaluación, se utilizarán técnicas de correspondencia de puntuación de la propensión para garantizar que las comunidades fuera del proyecto que fueron seleccionadas sean grupos de comparación adecuados. Se eliminará de la muestra cualquier control con atributos inusuales en relación con las comunidades de tratamiento (Anexo 1.15).

Fondo Hondureño de Inversión Social. El Fondo Hondureño de Inversión Social (FHIS) (véase Banco Mundial, 1998b) tiene el objetivo de mejorar las condiciones de vida de grupos sociales marginales mediante el financiamiento de subproyectos de infraestructura social y económica de pequeño tamaño. El FHIS es una institución guiada por la demanda que responde a las iniciativas de municipalidades, ministerios de gobierno, ONG y grupos de la comunidad. Proporciona financiamiento para inversiones en infraestructura, equipos y capacitación. La evaluación de impacto del FHIS emplea técnicas de comparación pareada, obteniendo aleatoriamente la muestra del grupo de tratamiento a partir de una lista de comunidades en que los proyectos del FHIS han estado en funcionamiento al menos un año. El grupo de comparación, en cambio, se seleccionó de una lista de proyectos en tramitación: aquellos que han sido solicitados o aprobados, pero para los cuales aún no se ha efectuado la inversión del FHIS. En teoría, las comunidades de los grupos de comparación se pareaan automáticamente con las comunidades del proyecto de acuerdo a un proceso de autoselección y los criterios de aprobación de proyectos del FHIS. Luego se realizó una encuesta domiciliaria en las comunidades del grupo de tratamiento y de comparación, complementada con un componente de evaluación cualitativa (grupos representativos y entrevistas a informantes clave) aplicado a un subconjunto de comunidades de tratamiento. La evaluación inicial es un primer paso hacia el establecimiento de un sistema supervisión y evaluación permanente que finalmente quede integrado a las operaciones del FHIS. Específicamente, los datos recolectados de las comunidades con proyectos en tramitación pasarán a ser una línea base útil con la cual se puede realizar un seguimiento de futuros cambios en los indicadores de efecto después que se aplique la inversión del FHIS.

Inversiones educacionales en la Región del Chaco de Bolivia. Los proyectos de educación financiados por el Fondo de Inversión Social (FIS) de Bolivia tienen el objetivo de mejorar las instalaciones físicas y capacitar a los maestros de escuelas públicas rurales. Los retrasos en la implementación del proyecto en la Región del Chaco y la limitación de fondos para el mejoramiento de las escuelas proporcionó la oportunidad de utilizar un diseño de evaluación experimental, asegurando al mismo tiempo que las escuelas más necesitadas se beneficiaran del proyecto. Las escuelas de la Región del Chaco se clasificaron de acuerdo con un índice de calidad de escuelas basado en la suma de cinco indicadores de infraestructura y equipos escolares: luz eléctrica, alcantarillado, fuente de agua, pupitres por alumno y metros cuadrados de espacio por alumno. Sólo las escuelas bajo un determinado valor límite estaban calificadas para recibir una intervención del FIS. Entre los establecimientos calificados, se seleccionaron automáticamente las escuelas en

peores condiciones para que se beneficiaran de las inversiones financiadas por el FIS. El grupo de la siguiente prioridad más alta incluía 120 escuelas, pero sólo se contaba con fondos para mejorar menos de la mitad de ellas. Así, las escuelas calificadas de este grupo de segunda prioridad se asignaron aleatoriamente a los grupos de tratamiento o de comparación, proporcionando las condiciones para un diseño de evaluación experimental (Anexo 1.4).

Combinación de métodos

Para la mayoría de las evaluaciones, se requiere más de una técnica para obtener resultados sólidos que respondan a varias preguntas de evaluación. Cada pregunta podría necesitar diferentes técnicas, incluso dentro de un único diseño de proyecto. Tres ejemplos ilustran la forma en que se combinaron diversas técnicas en una sola evaluación; el Fondo Social de Bolivia, la Evaluación de TRABAJAR de Argentina y la Reforma Escolar de Nicaragua.

El Fondo Social de Bolivia. Los fondos sociales generalmente incluyen varios tipos de subproyectos y por consiguiente el diseño de una evaluación puede abarcar diversos métodos. En el fondo Social de Bolivia, el patrón de implementación del proyecto dictó los métodos de evaluación. En el caso de la educación, ya se habían identificado las escuelas que recibirían la intervención, por consiguiente, no se podía emplear la aleatorización. En su lugar, se adoptaron métodos de pareo. En el caso de los proyectos de salud, se usaron métodos reflexivos, porque la intervención se implementaría en todos los centros de salud de la región (véase el Anexo 1.4).

Uso de una variada combinación de componentes de evaluación: TRABAJAR II de Argentina. La evaluación de TRABAJAR incluye una serie de componentes diseñados para evaluar la forma en que el proyecto está logrando sus objetivos de política. El primer componente se basa en los datos de una encuesta domiciliaria para evaluar las ganancias de ingreso que obtuvieron los participantes de TRABAJAR. El segundo componente supervisa la asignación de fondos del programa (focalización), vigilando los cambios en el tiempo como resultado de la reforma. Este componente genera retroalimentación dos veces al año, la que se usa para perfeccionar la focalización del programa. Entre los componentes adicionales de la evaluación se incluye un análisis de costos y beneficios de los proyectos de infraestructura y evaluaciones sociales diseñadas para proporcionar retroalimentación de la comunidad sobre la implementación del proyecto. Cada uno de estos componentes fue aplicado dos veces. Se planean tres componentes futuros. Nuevamente se aplicará la técnica de investigación de comparación pareada, para evaluar el impacto de la participación en el programa TRABAJAR sobre la actividad del mercado laboral. Se volverá a evaluar la calidad del proyecto de infraestructura, esta vez para los proyectos que fueron completados al menos un año atrás, para evaluar la durabilidad, mantenimiento y tasas de utilización. Por último, un componente cualitativo de investigación permitirá investigar las operaciones y procedimientos del programa, entrevistando a miembros del personal de los organismos que patrocinan proyectos y a beneficiarios del programa.

Los resultados de la evaluación proporcionan una clara orientación para la reforma de

políticas. El primer componente de evaluación revela que el programa TRABAJAR es muy exitoso en llegar a los pobres: la autoselección de los participantes mediante el ofrecimiento de bajos salarios es una estrategia que funciona en Argentina. Además, los participantes sí experimentan ganancias en su ingreso como resultado de la participación. El segundo componente permite concluir que la asignación geográfica de los fondos del programa ha mejorado en el tiempo; ahora el programa consigue dirigir mejor los fondos hacia las áreas pobres. Sin embargo, el proceso de evaluación permanente indica que el desempeño no es parejo y que incluso es persistentemente débil en algunas pocas provincias, conclusiones a las cuales en la actualidad se está prestando mayor atención de parte de las políticas. Los decepcionantes resultados de evaluación sobre la calidad de los proyectos de infraestructura han generado enormes efectos por parte del equipo del proyecto para mejorar los resultados en esta área a través de reformas de política. Se está insistiendo en realizar más visitas a terreno para evaluación y supervisión, sancionar a los organismos con resultados deficientes al finalizar el proyecto y fortalecer el manual de evaluación. Y por último, las evaluaciones sociales revelaron que es necesario prestar una mejor asistencia técnica a las ONG y municipalidades rurales durante la preparación e implementación de los proyectos, además de dar mayor publicidad y transparencia a la información sobre el programa TRABAJAR (Anexo 1.1).

Reforma de autonomía escolar de Nicaragua. En 1993, el gobierno nicaragüense adoptó medidas decisivas para implementar un importante esfuerzo de descentralización en el sector de la educación, otorgando autonomía presupuestaria y administrativa a un grupo de escuelas primarias y secundarias seleccionadas. La meta de las reformas es mejorar el aprendizaje de los alumnos: a medida que la administración de las escuelas sea más democrática y participativa, los patrones administrativos y de gasto de las escuelas locales se pueden orientar hacia esfuerzos que mejoren directamente la calidad pedagógica e impulsen el rendimiento de los alumnos. El efecto de esta reforma se ha evaluado usando una combinación de técnicas cuantitativas y cualitativas para examinar los resultados y el proceso de descentralización. El propósito del componente cualitativo es aclarar si en las escuelas se están efectivamente realizando las reformas administrativas y financieras previstas, y por qué y por qué no. El componente cuantitativo refuerza estos resultados al responder a la pregunta: "¿Los cambios administrativos y financieros en las escuelas realmente mejoran los resultados de aprendizaje en los niños?" Los resultados cualitativos indican que una implementación exitosa de las reformas depende en gran medida del marco y del entorno (es decir, el nivel de pobreza de la comunidad), mientras que los resultados cuantitativos indican que la mayor independencia en la toma de decisiones de las escuelas está asociado en forma significativa con un mejor rendimiento de los alumnos.

Se utilizaron metodologías y fuentes de datos diferentes, pero complementarias, para combinar con los enfoques. Por una parte, la evaluación cuantitativa siguió un diseño cuasi experimental, en el cual se comparan las calificaciones de las pruebas de una muestra de alumnos de escuelas autónomas (grupo de tratamiento) con los resultados de una muestra pareada de escuelas públicas no autónomas y escuelas privadas (grupo de comparación). Los datos para este componente de la evaluación se recopilaron de un panel de dos encuestas pareadas de escuelas-hogares y de las calificaciones de pruebas de rendimiento de

los alumnos. Por otra parte, el diseño de evaluación cualitativo consistía en una serie de entrevistas a informantes clave y conversaciones con grupos representativos con diferentes padres y miembros del personal de la escuela de una submuestra de escuelas autónomas y tradicionales incluida en el estudio cuantitativo.

Al usar técnicas de investigación cualitativas y cuantitativas, se generó una valiosa combinación de resultados útiles y pertinentes para las políticas. El trabajo cuantitativo proporcionó una amplia visión general y estadísticamente válida de las condiciones y resultados de las escuelas; en el trabajo cualitativo se perfeccionaron estas conclusiones con información sobre por qué algunos resultados esperados del programa de reforma han sido exitosos, mientras que otros han fracasado, lo que sirvió de base para los ajustes de la política. Además, como el trabajo cualitativo es más intuitivo, era más accesible y, por consiguiente, más interesante para el personal del Ministerio, lo que a su vez facilitó una rápida creación de capacidades para el proceso de evaluación dentro del Ministerio, aumentando su credibilidad (Anexo 1.11).

Aprovechamiento de las fuentes de datos existentes

Las fuentes de datos existentes, como una encuesta domiciliaria nacional, censo, registros administrativos de un programa o datos municipales, pueden aportar valiosa información a los esfuerzos de evaluación. Aprovechar las fuentes existentes reduce la necesidad de recurrir a una costosa recopilación de datos para el exclusivo propósito de una evaluación, como se ilustra en el caso de la evaluación de Caminos Rurales de Vietnam. Además, aunque los datos existentes podrían no contener toda la información que idealmente se reuniría para los fines de la evaluación, hay técnicas de evaluación innovadoras que con frecuencia pueden compensar los datos faltantes, como se muestra en el Proyecto de Extensión Agrícola Nacional de Kenya.

El Proyecto de caminos rurales de Vietnam. Los datos que se usaron en esta evaluación se basaron en una combinación eficaz de los datos nacionales y locales existentes con las encuestas realizadas específicamente para los fines de la evaluación. La encuesta domiciliaria de la evaluación fue diseñada con eficiencia para repetir varias preguntas de la Encuesta Nacional sobre las Condiciones de Vida de Vietnam, de modo que, al basarse en información que es común para ambas encuestas, se pueden usar técnicas de regresión para estimar la posición de cada hogar en la distribución nacional del bienestar.

La evaluación se basa en gran medida en los datos a nivel de comunal recopilados anualmente por las comunas y que cubren aspectos demográficos, uso de la tierra y actividades productivas. Esta fuente de datos se ve aumentada con una encuesta a nivel comunal realizada para la evaluación. Se crearon dos bases de datos adicionales usando la información existente. Se estableció una extensa base de datos a nivel provincial para ayudar a comprender la selección de las provincias dentro del proyecto. Esta base de datos abarca todas las provincias de Vietnam y los datos cubren un gran número de variables socioeconómicas. Por último, también se creó una base de datos a nivel de proyecto para cada una de las áreas del proyecto en que se realizaron encuestas, para controlar por la

magnitud del proyecto y el método de implementación para evaluar el impacto del proyecto (Anexo 1.15).

Proyecto de Extensión Nacional de Kenya (NEP). Los resultados del Proyecto de Extensión Nacional de Kenya (NEP) han sido polémicos y forman parte de un debate mayor sobre la eficacia en función de los costos del enfoque de capacitación y visitas (T&V) frente a los servicios de extensión agrícolas. En el contexto de Kenya, se ha intensificado el debate debido, por una parte, los altísimos rendimientos estimados para la capacitación y visitas en un estudio (Bindlish, Evenson y Gbetibouo, 1993, 1997) y, por otra, la falta de resultados visibles convincentes, incluido el rendimiento deficiente de la agricultura de Kenya en los últimos años.

El desacuerdo con respecto a los resultados del Proyecto de Extensión Nacional de Kenya se mantiene en espera de los resultados de esta evaluación, la cual se diseñó para asumir un enfoque riguroso y empírico frente a la evaluación del desarrollo institucional del programa y su efecto en el rendimiento agrícola. En la evaluación se utiliza una combinación de métodos cualitativos y cuantitativos para plantear preguntas sumamente pertinentes para las políticas y revela serias falencias del programa: (a) el desarrollo institucional del NEP ha sido limitado y después de 15 años hay pocos progresos en la eficacia de sus servicios; (b) la calidad y cantidad de los servicios prestados es deficiente y (c) los servicios de extensión sólo tienen un pequeño efecto positivo en la eficiencia agrícola y ninguno en la productividad de las explotaciones agrícolas.

De los resultados de la evaluación se puede obtener una serie de conclusiones concretas para las políticas, muchos de las cuales son pertinentes para el diseño de futuros proyectos de extensión agrícola. En primer lugar, la evaluación revela la necesidad de mejorar la focalización de la capacitación y visitas, centrándose en las áreas y grupos en que el efecto probablemente sea mayor. Además, la asesoría se debe adaptar cuidadosamente a las demandas de los agricultores, tomando en cuenta las variaciones de las condiciones tecnológicas y económicas locales. Para poder lograr este nivel de focalización del servicio, se requieren flujos adecuados de una información oportuna y confiable y, por consiguiente, un sistema de supervisión y evaluación del programa que genere un flujo constante de retroalimentación desde los beneficiarios sobre el contenido del servicio. Para aumentar la eficiencia del programa, es probable que una presencia más discreta y menos intensa de la capacitación y visitas, pero de cobertura más amplia, sea más eficaz en función de los costos. El enfoque general del programa frente a la entrega del servicio con una sola metodología (visitas a las explotaciones agrícolas) para entregar mensajes estándares también limita la eficiencia del programa. Es probable que con la reforma institucional mejore la eficacia en la entrega del servicio. La descentralización del diseño del programa, incluidos mecanismos de integración que hagan participar al agricultor (como participación en los costos y organizaciones de agricultores), debe llegar a ser una parte integral del mecanismo de entrega. Por último, la recuperación de costos, aun cuando sea sólo parcial, proporcionaría los incentivos adecuados, abordaría los temas de la rendición de cuentas y control de calidad, haría que el servicio se guiara más por la demanda y fuera más sensible y proporcionaría cierto mejoramiento presupuestario (Anexo 1.8).

Costos y financiamiento

Es indudable que una evaluación del impacto implica muchos costos, lo que explica por qué algunos países se muestran reticentes a financiar tales estudios. Estos costos incluyen la recopilación de datos y el valor del tiempo del personal para todos los miembros del equipo de evaluación. El financiamiento de una evaluación del impacto puede provenir de dentro del proyecto, de otros recursos gubernamentales, de una subvención para investigación o de un donante externo. La información sobre una muestra de evaluaciones del Banco Mundial indica que aunque muchos países asumieron la mayor parte de los costos de la evaluación, la implementación exitosa de una evaluación de impacto requirió considerables recursos externos más allá de los previstos en un préstamo o crédito para un proyecto. Estos recursos provinieron de una combinación de las siguientes fuentes: (a) préstamo o financiamiento de crédito del Banco Mundial para la recopilación y el procesamiento de datos; (b) el gobierno, a través de los sueldos pagados al personal local asignado al esfuerzo de evaluación (como se explica en el cuadro 4.1, estos costos de personal no se incluyeron en el cálculo de los costos de las evaluaciones realizadas para los casos que aquí se examinan, debido a las limitaciones de datos.); (c) subvenciones para investigación del Banco Mundial y subvenciones de donantes bilaterales que financiaron la asistencia técnica de asesores con los conocimientos especializados específicos requeridos para la evaluación y (d) el presupuesto de gastos generales del Banco Mundial, a través del tiempo del personal proporcionado para guiar la evaluación de impacto y que con frecuencia participa activamente en el trabajo analítico.

Aunque en pocas evaluaciones de impacto se documenta el costo de realizar el trabajo, en el Cuadro 4.2 se proporciona estimaciones de costos para una muestra de evaluaciones de impacto en que participó el Banco Mundial. Sin embargo, estas estimaciones de costos no incluyen el valor del tiempo del personal que aportaron al proceso las contrapartidas del país cliente (que puede ser significativo), porque no se contaba con esta información. Como referencia, en los 8 casos anteriores, fue habitual que hasta cinco miembros del personal estuvieran asignados al esfuerzo de evaluación durante varios años, un nivel de esfuerzo suficiente como para elevar significativamente el costo de la evaluación en muchos de los casos.

Cuadro 4.2 Resumen de costos estimados de diversas evaluaciones de impacto del Banco Mundial

Proyecto	Costo estimado de la evaluación (US\$) ^a	Costo como % del costo total del proyecto ^b	Costo como % del préstamo del BIRF o crédito de la AIF ^b	Desglose de los costos de evaluación (%)			
				Viaje ^c	Personal del Banco Mundial	Asesores	Recopilación de datos
Nicaragua Administración basada en las escuelas	495.000	1,26%	1,5%	8,1%	18,1%	39,0%	34,8%
El Salvador Administración basada en las escuelas	443.000	0,60%	1,3%	7,7%	7,4%	25,8%	59,2%
Colombia Programa de vales	266.000	0,20%	0,3%	9,4%	9,8%	21,8%	59,0%
Honduras Fondo Social	263.000	0,23%	0,9%	3,0%	11,5%	53,2%	32,3%
Nicaragua Fondo Social	449.000	0,30%	0,8%	4,9%	33,0%	7,8%	55,7%
Bolivia Fondo Social	878.000	0,50%	1,4%	3,4%	14,6%	12,9%	69,1%
Trinidad y Tobago Capacitación de Jóvenes	238.000	0,80%	1,2%	7,6%	11,5%	17,9%	63,1%
PROMEDIO	433.000	0,56%	1,0%	6,3%	15,1%	25,5%	53,3%

a. Este costo no incluye el costo del personal local de contrapartida no financiados con el préstamo o crédito. Las cifras se refieren al período dentro del cual se seleccionaron los proyectos de la muestra de evaluaciones y no al financiamiento total proporcionado alguna vez por el Banco y otros a esas instituciones.

b. Estos costos, como porcentaje del préstamo o crédito o del proyecto, se presentan sólo a modo de referencia. En muchos casos, el financiamiento real de la evaluación se obtuvo de fuentes ajenas al financiamiento del proyecto.

c. Las estimaciones de costo de viajes incluyen el viaje de la misión del personal del Banco Mundial y de asesores internacionales a los países clientes, al igual que viajes de contrapartidas del país cliente, en especial para participar en las sesiones de estrategia y talleres analíticos con asesores internacionales y personal del Banco Mundial.

Fuente: Archivos de Proyectos del Banco Mundial.

El costo estimado promedio para la evaluación del impacto fue de US\$433.000. Esto refleja una gama de US\$263.000 para la evaluación de un programa de capacitación en habilidades vocacionales para jóvenes desempleados de Trinidad y Tobago y US\$878.000 para la evaluación del Fondo de Inversión Social de Bolivia. El gasto en las evaluaciones de impacto para los proyectos a continuación refleja, en promedio, un 0,6% del costo total del proyecto (que a veces incluía financiamiento de varios donantes) o un 1,3% del costo del préstamo del Banco Internacional de Reconstrucción y Fomento (BIRF) o del crédito de la Asociación Internacional de Fomento (AIF). Los componentes más costosos de las

evaluaciones indicadas a continuación fueron la recopilación de datos y los asesores locales e internacionales. En muchos de los casos, los costos de viajes incluyeron viajes del personal local para reunirse con personal del Banco Mundial e investigadores para sesiones de estrategia y capacitación, porque la creación de capacidades entre el personal del país cliente constituía un objetivo clave. Los dos ejemplos siguientes para las evaluaciones de impacto de proyectos en Trinidad y Tobago y Bolivia ilustran algunas de las fluctuaciones que pueden surgir en los costos de los programas.

La evaluación del programa de capacitación en habilidades vocacionales de Trinidad y Tobago aprovechó una encuesta nacional sobre ingreso y empleo para tomar muestras de los graduados del programa y crear un grupo de comparación de un subconjunto de la muestra nacional. Además, el equipo de evaluación ayudó a diseñar y usar los datos administrativos disponibles de registros de postulantes al programa, de modo que se contaba con datos previos a la intervención y no se requirió empadronamiento para ubicar a los graduados del programa. El tamaño total de la muestra para cada uno de los tres estudios de seguimiento fue de aproximadamente 2500 jóvenes, incluidos los grupos de tratamiento y de comparación. En la encuesta sólo se aplicó un breve cuestionario y sólo a los graduados del programa. Trinidad y Tobago es un país pequeño, las comunidades son relativamente accesibles por tierra y el idioma común es el inglés en el país y entre los graduados del programa.

La evaluación del Fondo Social de Bolivia (FIS) utilizó sus propias encuestas básicas y de seguimiento de los grupos de tratamiento y comparación para evaluar las intervenciones en salud, educación, agua y servicios sanitarios. No había encuestas nacionales disponibles con las cuales realizar los análisis o efectuar sobremuestreos y por ello la evaluación se llevó toda la carga de la recopilación de datos. La muestra de los grupos de tratamiento y de comparación se componía de 7.000 hogares y 300 establecimientos entrevistados en la encuesta básica de 1993 y en la encuesta de seguimiento de 1998.

En Bolivia, los instrumentos de recopilación de datos para la evaluación del impacto incluyeron laboratorios portátiles para realizar pruebas de calidad del agua en laboratorio, pruebas de rendimiento y ocho cuestionarios para declarantes de los hogares y establecimientos. Los ocho cuestionarios constaban de dos cuestionarios domiciliarios (uno para el declarante principal y uno para mujeres en edad fértil), un cuestionario para la comunidad, cuatro diferentes cuestionarios de centros de salud para los distintos tipos de éstos (los que fluctúan desde pequeños consultorios locales hasta hospitales) y un cuestionario para escuelas, para el director de éstas. Para evaluar la focalización, la evaluación incluyó una medida de la pobreza basada en el consumo, lo que planteó la necesidad de recopilar datos detallados sobre el consumo de los hogares, como asimismo datos regionales sobre precios de las comunidades. El trabajo en terreno se realizó en áreas rurales, donde se ubica la mayoría de los proyectos del FIS, e incluyó una muestra aleatoria de hogares rurales a los cuales con frecuencia sólo se podía acceder a pie o a caballo. Por último, los cuestionarios tuvieron que prepararse y aplicar en español, quechua y aymará.

Asuntos de economía política

Existen varios aspectos de política económica que influyen no sólo en si se lleva a cabo una evaluación, sino también en la forma en que se implementa. La decisión de realizar una evaluación depende mucho de un sólido apoyo político. Muchos gobiernos no aprecian el valor de efectuar proyectos de evaluación y por lo tanto no desean invertir recursos en ellos. Además puede haber reticencia a permitir la ejecución de una evaluación independiente que pueda revelar resultados contrarios a la política del gobierno, especialmente en regímenes autoritarios o cerrados. Sin embargo, los gobiernos más abiertos podrían considerar que las evaluaciones y la divulgación de los resultados forman parte importante del proceso democrático.

Las evaluaciones también son sensibles al cambio político. Tres de los ocho estudios de efectos indicados en el Cuadro 4.2 se cancelaron debido a problemas de economía política. Las rotaciones de regímenes o de puestos clave dentro de un departamento gubernamental de contrapartida y los cambios en estrategias de política pueden afectar no sólo el esfuerzo de evaluación, sino más fundamentalmente la implementación del programa que se está evaluando. Un ejemplo de este tipo de riesgo lo experimentó un equipo que trabajaba en el diseño y evaluación del impacto de un programa piloto de administración basada en escuelas del Perú, como parte de un proceso de educación primaria financiado por el Banco Mundial. El equipo, compuesto por funcionarios del Ministerio de Educación, personal del Banco Mundial y asesores internacionales y locales, había trabajado por más de un año en la elaboración de los modelos de administración basada en escuelas que se probarían, en el establecimiento de un diseño experimental, en el diseño de instrumentos de encuestas y pruebas de rendimiento y en la recopilación de datos básicos sobre las características de las escuelas y el rendimiento escolar. Justo antes de la introducción de la prueba piloto en las escuelas seleccionadas en forma aleatoria, funcionarios gubernamentales de alto nivel cancelaron el experimento de administración basada en las escuelas en una reacción ante las repercusiones políticas percibidas del programa piloto. Varios años después se introdujo una reforma similar, pero sin el beneficio de una prueba piloto ni una evaluación.

En Venezuela, la evaluación de un programa de salud y nutrición materno-infantil se rediseñó tres veces con tres contrapartidas diferentes del cliente a medida que el gobierno transfería la responsabilidad de la evaluación de un organismo a otro. Cada cambio estuvo acompañado de una renegociación de contrato con la empresa del sector privado que se había identificado para realizar la recolección de datos y la mayor parte de los análisis para la evaluación. Cuando la legitimidad de la tercera contrapartida del gobierno comenzó a ser cuestionada, la empresa anuló el contrato y se abandonó la evaluación. Estos incidentes se produjeron durante un período de cambio político caracterizado por varias reorganizaciones de gabinete, las que terminaron con el fracaso del gobierno elegido que servía como contrapartida del proyecto. De modo que la evaluación difícilmente era la única que se veía afectada por las repercusiones de la inestabilidad política. Sin embargo, en los casos de Perú y Venezuela, es sensato reflexionar sobre la cantidad de recursos destinados a un esfuerzo que nunca se materializó. Un ejemplo menos dramático del efecto del cambio político en las estrategias de evaluación proviene de El Salvador, donde el éxito reconocido

de una reforma introducida en escuelas rurales hizo que el gobierno introdujera una reforma educacional similar en todas las escuelas urbanas de una sola vez, en lugar de introducirla gradualmente en las escuelas en el lapso de tiempo planificado originalmente. Esta decisión no permitió usar un diseño experimental y sólo se pudo usar una comparación reflexiva menos sólida como la única opción de diseño viable.

Bibliografía

La palabra *procesado* es una descripción de los trabajos reproducidos informalmente que podrían no estar disponibles habitualmente en bibliotecas.

Abadie, A., J. Angrist y G. Imbens. 1998. "Instrumental Variables Estimation of Quartile Treatment Effects." National Bureau of Economic Research, Serie de documentos de trabajo, N° 229:1–28.

Atkinson, Anthony. 1987. "On the Measurement of Poverty." *Econometrica* 55: 749–64.

Bamberger, Michael. 2000. *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C. Banco Mundial

Barnes, Carolyn. 1966. "Assets and the Impact of Microenterprise Finance Programs." USAID AIMS Project Brief 6. Washington, D.C.: USAID

Barnes, Carolyn, y Erica Keogh. 1999. "An Assessment of the Impact of Zambuko's Microenterprise Program in Zimbabwe: Baseline Findings." USAID AIMS Project Brief 23. Washington, D.C.: USAID

Benus, Jacob, Neelima Grover y Recep Varcin. 1998. *Turkey: Impact of Active Labor Programs*. Bethesda, MD: Abt Associates.

Benus, Jacob, Neelima Grover, Jiri Berkovsky y Jan Rehak. 1998. *Czech Republic: Impact of Active Labor Programs*. Bethesda, MD: Abt Associates.

Besharov, Douglas J., Peter Germanis y Peter H. Rossi. 1997. *Evaluating Welfare Reform: A Guide for Scholars and Practitioners*. College Park, Md.: University of Maryland.

Bindlish, V., R. Evenson y Mathurin Gbetibouo. 1993. *Evaluation of T & V-Based Extension in Burkina Faso*. Washington, D.C. Banco Mundial

Bloom, Howard S., Larry L. Orr, George Cave, Stephen H. Bell, Fred Doolittle y Winston Lin. 1994. *The National JTPA Study: Overview, Impacts, Benefits, and Costs of Title II-A*. Bethesda, Md.: Abt Associates.

Blundell, Richard W., y R. J. Smith. 1993. "Simultaneous Microeconomic Models with Censoring or Qualitative Dependent Variables." En G. S. Maddala, C. R. Rao y H. D. Vinod, eds., *Handbook of Statistics*. Vol. 11. Amsterdam: Holanda del Norte.

Bourguignon, Francois, Jaime de Melo y Akiko Suwa. 1991. "Distributional Effects of Adjustment Policies: Simulations for Archetype Economies in Africa and Latin America." *World Bank Economic Review* 5 (2): 339–66.

- Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9 (Primavera): 63–84.
- Card, David, y Philip K. Robins. 1996. "Do Financial Incentives Encourage Welfare Recipients to Work? Evidence from a Randomized Evaluation of the Self-Sufficiency Project." National Bureau of Economic, Documento de estudio 5701, agosto. Cambridge, Mass.: NBER.
- Carvalho, Soniya, y Howard White. 1994. "Indicators for Monitoring Poverty Reduction." Documentos de trabajo 254 del Banco Mundial. Washington, D.C.: Banco Mundial
- Chen, Martha Alter, y Donald Snodgrass. 1999. "An Assessment of the Impact of Sewa Bank in India: Baseline Findings." USAID AIMS Project, procesado. Washington, D.C. USAID
- Cohen, Monique, y Gary Gaile. 1997. "Highlights and Recommendations of the Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies." USAID AIMS Project. Washington, D.C.: Management Systems International.
- . 1998. "Highlights and Recommendations of the Second Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies." USAID AIMS Project. Washington, D.C.: Management Systems International.
- Dar, Amit, e Indermit S. Gill. 1998. "Evaluating Retraining Programs in OECD Countries: Lessons Learned." *The World Bank Research Observer* 13 (febrero): 79–101.
- Dar, Amit, y Zafiris Tzannatos. 1999. "Active Labor Market Programs: A Review of Evidence from Evaluations." Documento de trabajo 9901 de protección social del Banco Mundial. Washington, D.C.: Banco Mundial
- Dehejia, Rajeev H., y Sadek Wahba. 1998. "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." NBER, Serie de documentos de trabajo 6586. Cambridge, Mass.: NBER <http://www.nber.org/papers/w6586>.
- Dennis, Michael L., y Robert F. Boruch. 1989. "Randomized Experiments for Planning and Testing Projects in Developing Countries: Threshold Conditions." *Evaluation Review* 13 (junio): 292–309.
- Diagne, Aliou, y Manfred Zeller. 1998. "Determinants of Access to Credit and Its Impact on Income and Food Security in Malawi." Manuscrito presentado al Comité de revisión de publicaciones del Instituto Internacional de Investigaciones sobre Políticas Alimentarias (IFPRI) para su consideración como un Informe de Investigación IFPRI.
- Diop, F., A. Yazbeck, y R. Bitran. 1995. "The Impact of Alternative Cost Recovery

- Schemes on Access and Equity in Niger.” *Health Policy and Planning* 10 (3): 223–40.
- Donecker, Jane, y Michael Green. 1998. *Impact Assessment in Multilateral Development Institutions*. Londres: Department for International Development.
- Dunn, Elizabeth. 1999. “Microfinance Clients in Lima, Peru: Baseline Report for AIMS Core Impact Assessment.” USAID AIMS Project, procesado. Washington, D.C.: USAID
- Edgecomb, Elaine L. y Carter Garber. 1998. “Practitioner-Led Impact Assessment: A Test in Honduras.” USAID AIMS. Washington, D.C.: USAID.
- Ezemenari, Kene, Anders Ruqvist y K. Subbarao. 1999. “Impact Evaluation: A Note on Concepts and Methods.” World Bank Poverty Reduction and Economic Management Network, procesado. Washington, D.C.: Banco Mundial.
- Foster, James, J. Greer y Erik Thorbecke. 1984. “A Class of Decomposable Poverty Measures.” *Econometrica* 52: 761–65.
- Friedlander, Daniel, y Gayle Hamilton. 1993. *The Saturation Work Initiative Model in San Diego: A Five-Year Follow-Up Study*. Nueva York: Manpower Demonstration Research Corporation.
- Friedlander, Daniel, y Philip K. Robins. 1995. “Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods.” *American Economic Review* 85 (septiembre): 923–37.
- Friedlander, Daniel, David H. Greenberg y Philip K. Robins. 1997. “Evaluating Government Training Programs for the Economically Disadvantaged.” *Journal of Economic Literature* 35 (diciembre): 1809–55.
- Fuller, Bruce, y Magdalena Rivarola. 1998. *Nicaragua’s Experiment to Decentralize Schools: Views of Parents, Teachers and Directors*. Serie de documentos de trabajo sobre Evaluación de reformas de la educación, documento N° 5. Washington, D.C.: Banco Mundial.
- Gaile, Gary L., y Jenifer Foster. 1996. “Review of Methodological Approaches to the Study of the Impact of MicroEnterprise Credit Programs.” USAID AIMS Brief 2. Washington, D.C.: USAID
- Glewwe, Paul, Michael Kremer y Sylvie Moulin. 1998. “Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya.” Development Research Group, Banco Mundial, 28 de octubre
- Goodman, Margaret, Samuel Morley, Gabriel Siri y Elaine Zuckerman. 1997. *Social Investment Funds in Latin America: Past Performance and Future Role*. Washington,

D.C. Banco Interamericano de Desarrollo. Marzo.

Gobierno de Dinamarca. 1995. *Methods for Evaluation of Poverty Oriented Aid Interventions*. Copenhague: Ministerio de Asuntos Exteriores.

Greene, W. H. 1997. *Econometric Analysis*. Hemel Hempstead, New Jersey: Prentice Hall Press.

Greenberg, David, y Mark Shroder. 1997. *The Digest of Social Experiments*, segunda edición. Washington, D.C.: The Urban Institute Press.

Grosh, Margaret E., y Juan Muñoz. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey." LSMS, Documento de trabajo N° 126. Banco Mundial, Washington, D.C.

Grossman, Jean Baldwin. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *The World Bank Research Observer* 9 (julio): 159–80.

Habicht, J. P., C. G. Victoria y J. P. Vaughan. 1999. "Evaluation Designs for Adequacy, Plausibility and Probability of Public Health Programme Performance and Impact." *International Journal of Epidemiology* 28: 10–18.

Harrell, Adele, Martha Burt, Harry Hatry, Shelli Rossman, Jeffrey Roth y William Sabol. 1996. *Evaluation Strategies for Human Service Programs: A Guide for Policymakers and Providers*. Washington, D.C.: Urban Institute.

Heckman, James. 1997. "Instrumental Variables. A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–61.

Heckman, James, R. Lalonde y J. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." En O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. III. Nueva York: Elsevier Science Publishing Co.

Heckman, James, y Richard Robb. 1985. "Alternative Methods of Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30: 239–67.

Heckman, James J., y Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9 (2): 85–110.

Heckman, J., H. Ichimura, J. Smith y P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66: 1017–99.

Hicks, Norman. 1998. *Measuring the Poverty Impact of Projects in LAC*. Región de América Latina y el Caribe del Banco Mundial, procesado. Washington, D.C.:

Banco Mundial.

Holder, Harold D., Andrew J. Treno, Robert F. Saltz y Joel W. Grube. 1997. "Summing Up: Recommendations and Experiences for Evaluation of Community-Level Prevention Programs." *Evaluation Review* 21 (abril): 268–78.

Hollister, Robinson G., y Jennifer Hill. 1995. *Problems in the Evaluation of Community-Wide Initiatives*. Nueva York: Russell Sage Foundation.

Hulme, David. 1997. "Impact Assessment Methodologies for Microfinance: Theory, Experience and Better Practice." Manchester, R.U.: Institute for Development Policy and Management, University of Manchester.

Instituto Internacional de Investigaciones sobre Políticas Alimentarias. 1998. *Programa Nacional de Educación, Salud, y Alimentación (PROGRESA): A Proposal for Evaluation* (con apéndice técnico). Washington, D.C.: IFPRI.

Jalan, Jyotsna, y Martin Ravallion. 1998a. "Income Gains from Workfare: Estimates for Argentina's TRABAJAR Program Using Matching Methods." Washington, D.C. Development Research Group, Banco Mundial.

———. 1998b. "Transfer Benefits from Workfare: A Matching Estimate for Argentina." Washington, D.C.: Development Research Group, Banco Mundial.

Jamison, Dean T., Barbara Serle, Klaus Galda y Stephen P. Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement." *Journal of Educational Psychology* 73 (4): 556–67.

Karoly, Lynn A., Peter W. Greenwood, Susan S. Everingham, Jill Houbé, M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders y James Chiesa. 1998. *Investing in Our Children: What We Know and Don't Know about the Costs and Benefits of Early Childhood Interventions*. Santa Monica, Calif.: Rand.

Kemple, James J., Fred Doolittle y John W. Wallace. 1993. *The National JTPA Study: Site Characteristics and Participation Patterns*. Nueva York: Manpower Demonstration Research Corporation.

Khandker, Shahidur R. 1998. *Fighting Poverty with Microcredit: Experience in Bangladesh*. Nueva York: Oxford University Press para el Banco Mundial.

Killick, Tony. 1995. *IMF Programmes in Developing Countries, Design and Impact*. Londres: Inglaterra.

King, Elizabeth, y Berk Ozler. 1998. *What's Decentralization Got to Do with Learning?*

The Case of Nicaragua's School Autonomy Reform. Serie de documentos de trabajo sobre Evaluación de reformas de la educación, documento N° 9. Washington, D.C.: Banco Mundial.

King, Elizabeth, Berk Ozler y Laura Rawlings. 1999. *Nicaragua's School Autonomy Reform: Fact or Fiction?* Washington, D.C.: Banco Mundial.

Kish, Leslie. 1965. *Survey Sampling*. Nueva York: John Wiley and Sons.

Levinson, James F., Beatrice Lorge Rogers, Kristin M. Hicks, Thomas Schaetzel, Lisa Troy y Collette Young. 1998. *Monitoring and Evaluation: A Guidebook for Nutrition Project Managers in Developing Countries*. Medford, Mass.: International Food and Nutrition Center, Tufts University School of Nutrition Science and Policy, marzo.

Manski, Charles, e Irwin Garfinkel, eds. 1992. *Evaluating Welfare and Training Programs*. Cambridge, Mass.: Harvard University Press.

Manski, Charles, y Steven Lerman. 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica* 45: 1977–88.

McKay, H., A. McKay, L. Siniestra, H. Gomez y P. Lloreda. 1978. "Improving Cognitive Ability in Chronically Deprived Children" *Science* 200(21): 270–78.

Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics*, 13 (abril): 151–161.

Miles, Matthew B., y A. Michael Huberman. 1994. *Qualitative Data Analysis*. Londres: Sage Publications.

MkNelly, Barbara, y Karen Lippold. 1998. "Practitioner-Led Impact Assessment: A Test in Mali." USAID AIMS Brief 21. Washington, D.C.: USAID

MkNelly, Barbara, y Christopher Dunford (en colaboración con el Program in International Nutrition, University of California, Davis). 1998. "Impact of Credit with Education on Mothers' and Their Young Children's Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana." *Freedom from Hunger*, Documento de Estudio N° 4, marzo.

Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review* 15 (3): 291–314.

Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation*, segunda edición.

Thousand Oaks, Calif.: Sage Publications.

- . 1999. “The Qualitative Method of Impact Analysis.” Procesado.
- Morduch, Jonathan. 1998a. “The Microfinance Schism.” Harvard Institute for International Development, Documento de Análisis de Desarrollo 626, febrero.
- . 1998b. “Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh.” Harvard University Institute for International Development, procesado, junio.
- . 1999. “The Microfinance Promise.” *Journal of Economic Literature* 37(4): 1569–1614.
- Newman, John, Menno Pradhan, Laura Rawlings, Ramiro Coa y Jose Luis Evia. 2000. “An Impact Evaluation of Education, Health, and Water Supply Investments of the Bolivia Social Investment Fund.” Procesado.
- Newman, John, Laura Rawlings y Paul Gertler. 1994. “Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries.” *The World Bank Research Observer* 9 (Julio): 181–202.
- Operations Evaluation Department, Banco Mundial. 1994. “An Overview of Monitoring and Evaluation in the World Bank.” 30 de junio.
- Poppele, J. S. Summarto, y L. Pritchett. 1999. “Social Impacts of the Indonesia Crisis: New Data and Policy Implications.” Social Monitoring Early Response Unit, Banco Mundial, Washington, D.C. Procesado.
- Pradhan, Menno, Laura Rawlings y Geert Ridder. 1998. “The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation.” *World Bank Economic Review* 12 (3): 457–82.
- Ravallion, Martin. 1994. *Poverty Comparisons, Fundamentals in Pure and Applied Economics*, vol. 56. Harwood Academic Publishers.
- . 1999. *Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved*. Banco Mundial, Washington, D.C. Procesado.
- Ravallion, Martin, y Quentin Wodon. 1998. *Evaluating a Targeted Social Program When Placement Is Decentralized*. Documento de Trabajo de Estudio de Políticas del Banco Mundial, 1945. Washington, D.C.: Banco Mundial.
- Ravallion, Martin, Dominique van de Walle y Madhur Gautam. 1995. “Testing a Social Safety Net.” *Journal of Public Economics* 57: 175–99.
- Rawlings, Laura. 2000. “Assessing Educational Management and Quality in Nicaragua.” En

Bamberger: *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C.: Banco Mundial.

- Rebien, Claus C. 1997. "Development Assistance Evaluation and the Foundations of Program Evaluation." *Evaluation Review* 21 (agosto): 438–60.
- Revenga, Ana, Michelle Riboud y Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages." *World Bank Economic Review* 8 (2): 247–77.
- Rivers, Douglas, y Quang H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39: 347–66.
- Robinson, Sherman. 1991. "Macroeconomics, Financial Variables, and Computable General Equilibrium Models." *World Development* 19(11): 1509–25.
- Rosenbaum, P., y D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55.
- . 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *American Statistician* 39: 35–39.
- Rossmann, Gretchen B., y Bruce L. Wilson. 1985. "Numbers and Words: Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study." *Evaluation Review* 9 (3): 627–43.
- Sahn, David, Paul Dorosh y Stephen Younger. 1996. "Exchange Rate, Fiscal and Agricultural Policies in Africa: Does Adjustment Hurt the Poor?" *World Development* 24(4): 719–47.
- Sebstad, Jennifer, y Gregory Chen. 1996. "Overview of Studies on the Impact of MicroEnterprise Credit." USAID AIMS, junio. Washington, D.C.: USAID
- Sharma, Manohar, y Gertrud Schrieder. 1998. "Impact of Finance on Food Security and Poverty Alleviation—A Review and Synthesis of Empirical Evidence." Documento presentado en Workshop on Innovations in Micro-Finance for the Rural Poor: Exchange of Knowledge and Implications for Policy, Ghana, noviembre.
- Smith, William. 1998. *Group Based Assessment of Change: Method and Results*. Hanoi, Vietnam: ActionAid.
- Subbarao, K., Kene Ezemenari, John Randa y Gloria Rubio. 1999. *Impact Evaluation in FY98 Bank Projects: A Review*. World Bank Poverty Reduction and Economic Management Network, procesado, enero.

Tan, J.P., J. Lane y G. Lassibille. 1999. "Schooling Outcomes in Philippine Elementary Schools: Evaluation of the Impact of Four Experiments." *World Bank Economic Review*, septiembre.

Taschereau, Suzanne. 1998. *Evaluating the Impact of Training and Institutional Development Programs, a Collaborative Approach*. Economic Development Institute of the World Bank, enero.

USAID. 1984. "Study of Low Cost Housing." Nueva Delhi: USAID. Junio.

———. 1995. "Assessing the Impacts of MicroEnterprise Interventions: A Framework for Analysis." *MicroEnterprise Development Brief*, junio.

Valadez, J., y M. Bamberger, ed. 1994. "Monitoring and Evaluating Social Programs in Developing Countries." *Economic Development Institute of the World Bank Series*, Banco Mundial, Washington, D.C.

van de Walle, Dominique. 1999. "Assessing the Poverty Impact of Rural Road Projects." Banco Mundial. Procesado.

Walker, Ian, Rafael del Cid, Fidel Ordóñez y Florencia Rodríguez. 1999. "Evaluación Ex-Post del Fondo Hondureño de Inversión Social (FHIS 2)." *ESA Consultores*, Tegucigalpa, Honduras.

Weiss, Carol H. 1998. *Evaluation*. Upper Saddle River, New Jersey: Prentice Hall.

Wilson, Sandra Jo. 1998. "A Multi-Method Proposal to Evaluate the Impact of CGIAR-Instigated Activities on Poverty Reduction." Documento presentado en la Conferencia Anual de la American Evaluation Association, noviembre.

Banco Mundial, Poverty and Human Resources Division, Policy Research Department. 1998a. *Impact Evaluation of Education Projects Involving Decentralization and Privatization*. Documento de trabajo. Washington, D.C.: Banco Mundial.

———, Learning and Leadership Center and Operations Policy Department. 1996. *Handbook on Economic Analysis of Investment Operations*. Washington, D.C.: Banco Mundial.

———, Operations and Evaluation Department. 1998a. *1998 Annual Review of Development Effectiveness*. Washington, D.C.: Banco Mundial.

———, Poverty Reduction and Economic Management, Latin America and Caribbean Region. 1998b. *Measuring the Poverty Impact of Projects in LAC*. Washington, D.C.: Banco Mundial.

- , Operations and Evaluation Department. 1999. *World Bank Agricultural Extension Projects in Kenya: An Impact Evaluation*. Informe N° 19523. Washington, D.C.: Banco Mundial.
- , Poverty Reduction and Economic Management Unit, 1999. *Strategy for Impact Evaluation*. Washington, D.C.: Banco Mundial.
- Wouters, A. 1995. "Improving Quality through Cost Recovery in Niger." *Health Policy and Planning* 10 (3): 257–70.
- Zeller, Manfred, Akhter Ahmed, Suresh Babu, Sumiter Broca, Aliou Diagne y Manohar Sharma. 1996. "Security of the Poor: Methodologies for a Multicountry Research Project." IFPRI, Documento de trabajo 11.

Anexo 1

Estudios de casos

Anexo 1.1: Evaluación de los beneficios de los programas de trabajo público con salario mínimo para los pobres: El programa TRABAJAR de Argentina

I. Introducción

Descripción del proyecto. El objetivo del programa TRABAJAR de Argentina es disminuir la pobreza mediante la simultánea generación de oportunidades de empleo para los pobres y el mejoramiento de la infraestructura social en las comunidades pobres. TRABAJAR I, un programa piloto, se creó en 1996 en respuesta a una crisis económica de ese momento y a tasas de desempleo de más del 17%. TRABAJAR II fue lanzado en 1997 como una versión ampliada y reformada del programa piloto y los proyectos de TRABAJAR III se comenzaron a aprobar en 1998. En el programa se ofrecen salarios relativamente bajos, para así atraer (autoseleccionar) la participación sólo de trabajadores pobres y sin empleo. Los proyectos en que se contrata el trabajo de los participantes son propuestos por los gobiernos y organizaciones no gubernamentales (ONG) locales, los que deben cubrir los costos extrasalariales del proyecto. Los proyectos se aprueban en el nivel regional, de acuerdo a directrices que imparte el Gobierno Central.

Según el proceso de evaluación, el programa ha tenido cambios en su diseño y procedimientos operativos. En TRABAJAR II se aplicaron diversas reformas destinadas a mejorar la orientación de los proyectos hacia grupos beneficiarios específicos. Los indicadores de pobreza y desempleo provincial ahora tienen una mayor influencia en el sistema de asignación presupuestaria del Gobierno Central. Asimismo, las propuestas de proyectos provenientes de áreas pobres tienen una mayor ponderación dentro de las directrices de aprobación de los proyectos. En el nivel local, tanto en TRABAJAR II como TRABAJAR III se han hecho esfuerzos por fortalecer la capacidad de los ministerios locales para ayudar a las áreas pobres en el montaje de los proyectos y en elevar las normas de calidad de la infraestructura.

Evaluación del impacto. Los esfuerzos de evaluación se iniciaron durante la preparación de los proyectos para TRABAJAR II y siguen realizándose. Con la evaluación se busca determinar si el programa está cumpliendo sus metas de políticas, además de destacar las áreas donde se necesitan modificaciones para lograr un máximo de eficacia. La evaluación consta de varios estudios separados, con los cuales se evalúan a) las ganancias netas de ingreso que obtienen los participantes en el programa, b) la asignación de los

recursos del programa en las regiones (orientación hacia grupos de beneficiarios específicos), c) la calidad de los proyectos de infraestructura que se financiaron y d) el papel que desempeña la comunidad y las ONG en el resultado de los proyectos.

En cuanto a los aspectos técnicos, hay dos componentes de evaluación que se destacan en demostrar técnicas empíricas de prácticas óptimas. En primer lugar, el estudio de ganancias netas de ingreso muestra técnicas de prácticas óptimas en comparaciones pareadas, como asimismo un uso ingenioso de los datos de encuestas domiciliarias existentes para realizar el ejercicio de correspondencia. En segundo lugar, el estudio de la focalización de los resultados representa una nueva técnica para evaluar la orientación hacia beneficiarios específicos cuando no se observa la incidencia del gasto público en el nivel local. El diseño general de la evaluación también representa una práctica óptima de la combinación de componentes y técnicas de investigación (desde análisis cuantitativo pasando por visitas de diseño en el terreno hasta evaluaciones sociales), lo que presenta un flujo integral y oportuno de resultados.

II. Diseño de la evaluación

La evaluación de TRABAJAR incluye una serie de componentes diseñados para evaluar el grado en que el programa está cumpliendo con sus objetivos de políticas. El primer componente se basa en datos de encuestas domiciliarias para evaluar las ganancias de ingreso que obtienen los participantes de TRABAJAR. Al estimar sus ingresos netos, este estudio representa un mejoramiento con respecto a las evaluaciones convencionales de programas de trabajo público, en los que habitualmente se miden las ganancias de ingreso de los participantes tan sólo en términos de sus salarios *brutos* percibidos. Usando recientes avances en las técnicas de comparación pareada, el estudio da cuenta del ingreso sacrificado (el ingreso sacrificado por los participantes al integrarse al programa TRABAJAR), lo que arroja una estimación inferior y más exacta de las ganancias de ingreso para los participantes. Con el segundo componente se controla la asignación de fondos del programa (focalización) al examinar los cambios en el tiempo generados por la reforma. A través de un uso juicioso de los datos disponibles comúnmente (asignaciones de fondos del programa en las provincias y un censo nacional), el diseño de este componente representa una nueva metodología para evaluar la focalización en la pobreza cuando no existen datos reales sobre la incidencia del gasto en el marco del programa. Este análisis comenzó con la primera misión de supervisión (noviembre de 1997) y desde entonces se ha actualizado dos veces cada año.

Los componentes adicionales de la evaluación incluyen análisis de costos-beneficios realizados para una submuestra de proyectos de infraestructura, junto con evaluaciones sociales destinadas a entregar información sobre la puesta en práctica de los proyectos. Cada una de estas actividades se realizó dos veces, tanto para TRABAJAR II como para TRABAJAR III. Se han planificado tres futuras actividades de evaluación. La técnica de investigación de comparación pareada se aplicará nuevamente para evaluar la repercusión de la participación en el programa TRABAJAR sobre la actividad del mercado laboral. La calidad de los proyectos de infraestructura se volverá a evaluar esta vez para los

proyectos que finalizaron hace al menos un año, para evaluar su durabilidad, mantenimiento y tasas de utilización. Finalmente, en el marco de un componente de investigación cualitativa, se investigarán las operaciones y procedimientos del programa a través de entrevistas a miembros del personal de agencias que patrocinan los proyectos, como asimismo a sus beneficiarios.

III. Técnicas de recopilación y análisis de datos

La evaluación de las ganancias netas en ingreso para los participantes en el programa se basa en dos fuentes de datos, una encuesta de niveles de vida (la Encuesta de Desarrollo Social – EDS) y una encuesta de los participantes de TRABAJAR realizada específicamente para fines de evaluación (la encuesta EDS se financió en el marco de otro proyecto del Banco Mundial. En ella se buscaba mejorar la calidad de la información sobre el bienestar de los hogares en Argentina, especialmente en el área de acceso a los servicios sociales y los programas sociales públicos). Estas encuestas se llevaron a cabo en agosto (EDS) y septiembre (encuesta de participantes en TRABAJAR) de 1997 por la oficina nacional de estadísticas, usando el mismo cuestionario y los mismo equipos entrevistadores. La muestra de la encuesta EDS cubre un 85% de la población nacional, omitiendo algunas áreas rurales y comunidades muy pequeñas. La muestra de la encuesta de participantes en TRABAJAR se basa en una muestra aleatoria de los proyectos de TRABAJAR II ubicados dentro del marco de la muestra de EDS y genera datos para 2.802 participantes en los programas actuales (el total de los participantes en TRABAJAR II entre mayo del 97 y enero del 98 suma 65.321). La confiabilidad de la técnica de correspondencia mejora al poder aplicar simultáneamente el mismo cuestionario a los participantes y a los grupos de control y al poder asegurar que ambos grupos pertenezcan al mismo ambiente económico.

Para generar el grupo de control de correspondencia para la encuesta EDS, en el estudio se utiliza una técnica llamada puntuación de la propensión (el hecho de que el cuestionario de la EDS sea muy amplio y que se recopilen datos detallados sobre las características de los hogares que ayudan a predecir la participación en los programas facilita el uso de la técnica de puntuación de la propensión). Una correspondencia ideal sería de dos individuos, uno en la muestra de los participantes y uno en el grupo de control, para los cuales todas estas variables de (x) que predicen la participación en los programas fueran idénticas. El problema normal en una correspondencia es que esto es impracticable, debido al elevado número de variables incluidas en (x). Sin embargo, se pueden calcular correspondencias para la puntuación de propensión de cada individuo, lo que simplemente es la probabilidad de participar condicionada a (x) (con la ayuda de modelos logit estándar, se calcula la puntuación de la propensión para cada observación en la muestra de participantes y el grupo de control). Los datos sobre el ingreso en el grupo de control de correspondencia de no participantes permite estimar los ingresos sacrificados por los participantes efectivos en TRABAJAR II. Luego se calcula el ingreso neto que resulta de la participación en el programa como total de salarios del programa, menos el ingreso sacrificado.

El análisis de la orientación hacia beneficiarios específicos se destaca por no haber

requerido ninguna recopilación especial de información. El trabajo empírico se basa en datos de la oficina ministerial de proyectos sobre las asignaciones de fondos por departamento geográfico para TRABAJAR I (marzo de 1996-abril de 1997) y los primeros seis meses de TRABAJAR II (mayo a octubre de 1997). También se basa en un índice de pobreza para cada departamento (que son 510), calculado a partir del censo de 1991, como proporción de los hogares con “necesidades básicas no satisfechas. Se trata de un índice compuesto que representa el hacinamiento residencial, instalaciones sanitarias, calidad de la vivienda, nivel de instrucción de los adultos, matrícula escolar de los niños, empleo e índice de dependencia (relación miembros familiares que trabajan/no trabajan). El índice es algo antiguo, aunque esto tiene la ventaja de que la medida de la pobreza departamental es exógena a las intervenciones de TRABAJAR (no está influenciada por ellas). Para analizar la incidencia de la orientación hacia beneficiarios específicos, se hacen regresiones de los datos del gasto público por área geográfica (en este caso, por departamento) en los correspondientes índices de pobreza geográfica. El coeficiente resultante estima sistemáticamente un ‘diferencial de focalización’ dado por la diferencia entre las asignaciones promedio del programa a los pobres y los no pobres. Entonces se puede descomponer este diferencial nacional de la focalización para evaluar la contribución del mecanismo de focalización del Gobierno Central (asignaciones de fondos a través de los departamentos) con respecto a la focalización del gobierno local en el nivel provincial .

El análisis de costos-beneficios fue realizado por una ingeniero civil, la que efectuó un estudio de dos etapas de los proyectos de infraestructura de TRABAJAR. En la primera etapa, visitó una muestra de 50 proyectos de TRABAJAR I completados y los calificó sobre la base de indicadores en seis categorías: técnica, institucional, ambiental, socioeconómica, supervisión y operaciones y mantenimiento. Luego se calificó globalmente a los proyectos de acuerdo a un sistema de puntos y, donde correspondía, se realizaron análisis de costos-beneficios (no para escuelas y centros de salud). Un año después se realizó un estudio similar de seguimiento de 120 proyectos de TRABAJAR II, para evaluar el impacto de las reformas en la calidad de la infraestructura.

Durante la preparación de los proyectos se realizaron las evaluaciones sociales tanto para TRABAJAR I como TRABAJAR II. Estas evaluaciones entregan información sobre aspectos de la implementación de los proyectos, por ejemplo la función desempeñada por las ONG, la disponibilidad de asistencia técnica en la preparación y construcción de los proyectos y la selección de los beneficiarios. Ambas evaluaciones sociales fueron realizadas por sociólogos mediante grupos representativos y entrevistas.

IV. Resultados

Es importante considerar el ingreso sacrificado para lograr un cuadro preciso de los beneficios que otorgan los programas de trabajo público. Las estadísticas descriptivas de los participantes en TRABAJAR II parecen indicar que sin el acceso al programa (ingreso familiar per cápita menos los salarios del programa), cerca del 85% de los participantes en el programa habrían caído al 20% más bajo de la distribución nacional del ingreso, quedando calificados como pobres en Argentina. Sin embargo, las estimaciones del ingreso sacrificado que arroja el método de comparación son considerables, de modo que el ingreso neto promedio que se ganó mediante la participación en el programa es de

alrededor de la mitad del salario de TRABAJAR. Los participantes en Trabajar no tenían los recursos para estar sin empleo si el programa no hubiera existido. Por consiguiente, se sacrifica algún tipo de ingreso al participar en el programa. Es este ingreso sacrificado el que se estima al observar el ingreso de los que no participan en contraposición a los que participan en el programa. Sin embargo, incluso dando lugar al ingreso sacrificado, la distribución del beneficio es claramente pro pobres: un 80% de los participantes en el programa caen en el 20% más bajo de la distribución del ingreso. La participación de las mujeres es baja en el programa (15%), pero las ganancias netas de ingreso son prácticamente idénticas para los hombres y mujeres que participan en TRABAJAR; los participantes más jóvenes obtienen ganancias netas de ingreso significativamente más bajas.

El desempeño de la orientación hacia beneficiarios específicos mejoró marcadamente como resultado de las reformas de TRABAJAR II. La asignación implícita de los recursos a los hogares pobres en TRABAJAR I se multiplicó por siete en TRABAJAR II. Un tercio de este mejoramiento se debió a una mejor focalización en el nivel central, mientras que dos tercios fueron el resultado de una mejor focalización en el nivel provincial. Sin embargo, los resultados de la orientación hacia beneficiarios específicos difieren de manera importante entre las provincias. Un departamento con un 40% de sus habitantes clasificados como pobres puede esperar recibir entre cero a cinco veces la asignación departamental media, dependiendo de la provincia a la que pertenece. Asimismo, la tendencia era que el desempeño de esta focalización fuera peor en las provincias más pobres.

Se encontró que la calidad de los proyectos de infraestructura era adecuada; sin embargo, lo que es decepcionante, las reformas de TRABAJAR II no produjeron mejoramientos significativos. Esto se explica en parte por la aguda expansión del programa, lo que dificultó el logro de algunos de los niveles operativos que se habían especificado ex-ante. Sin embargo, sí hubo mayores logros en cuanto a satisfacer las necesidades de prioridad de las comunidades. La evaluación social reveló que se debía mejorar la asistencia técnica a las ONG y municipalidades rurales, además de dar una mayor publicidad y mejorar la transparencia de la información sobre el programa TRABAJAR.

V. Aplicación de políticas

Los resultados de la evaluación indican claramente que los participantes del programa TRABAJAR pertenecen en gran medida al sector de los pobres. La autoselección de los participantes mediante el ofrecimiento de salarios bajos es una estrategia que funciona en Argentina y los participantes se benefician en materia de ingresos como resultado de esta participación (aunque las ganancias netas son inferiores al salario bruto, debido al ingreso sacrificado). El programa no parece discriminar en contra de la participación femenina. Las reformas de TRABAJAR II han mejorado los resultados de la focalización geográfica (el programa ahora orienta mejor los fondos hacia las áreas pobres); sin embargo, hay diferentes niveles de desempeño y éste es persistentemente deficiente en una pocas provincias, donde se requieren políticas especiales. Finalmente, los decepcionantes resultados en la calidad de los proyectos de infraestructura han resultado en enormes esfuerzos del equipo del proyecto en cuanto a mejorar el desempeño en esta área

mediante el perfeccionamiento de los procedimientos de operación e insistiendo en más visitas a terreno para fines de evaluación y supervisión, castigando a las agencias que tienen un desempeño deficiente en la realización de los proyectos y fortaleciendo el manual de evaluación.

VI. Costos y administración de la evaluación

Costos. Los costos de implementar la encuesta de TRABAJAR (para el estudio de las ganancias netas de ingreso) y el procesamiento de la información sumaron aproximadamente US\$350.000. Las dos evaluaciones de la calidad de los subproyectos (análisis de costos-beneficios) costaron cerca de US\$10.000 cada una, al igual que las evaluaciones sociales, lo que eleva el gasto total estimado de la evaluación a US\$390.000.

Administración. La evaluación fue diseñada por Martin Ravallion, un miembro del personal del Banco Mundial, e implementada conjuntamente con el equipo del proyecto de Argentina y el Banco Mundial. En sus diversas etapas, durante la tarea de evaluación también fue necesaria la coordinación con varias agencias gubernamentales locales, entre ellas la agencia estadística, el Ministerio del Trabajo (incluidas las oficinas locales) y el departamento de análisis de políticas del Ministerio de Desarrollo Social.

VII. Lecciones aprendidas

La importancia de tener en cuenta el ingreso sacrificado al evaluar las ganancias de los trabajos públicos. El ingreso sacrificado representa una proporción importante (cerca de la mitad) del salario bruto percibido por los participantes en el programa de trabajo público en Argentina. Los resultados parecen indicar que los métodos de evaluación convencionales (usando sólo el salario bruto) sobrevaloran significativamente las ganancias de ingreso y por consiguiente, también sobrestiman cuán pobres serían los participantes en caso de no existir el programa.

El método de correspondencia de la puntuación a la propensión. Cuando se usa la técnica de evaluación de comparaciones pareadas, las puntuaciones de la propensión permiten establecer correspondencias confiables entre una muestra de participantes y no participantes (grupo de control).

Un uso sensato de las fuentes de datos nacionales existentes. Con frecuencia, las fuentes de datos existentes como un censo nacional o una encuesta domiciliaria, pueden ser un aporte valioso a los esfuerzos de evaluación. Al basarse en fuentes existentes, disminuye la necesidad de recurrir a una costosa recopilación de datos sólo para fines de evaluación. Hay técnicas de evaluación ingeniosas que pueden compensar la falta de datos, como lo ha demostrado adecuadamente la evaluación de los resultados de la focalización geográfica de TRABAJAR.

Amplia gama de componentes de evaluación. El diseño de la evaluación de TRABAJAR es el ejemplo de una combinación efectiva de herramientas y técnicas de

evaluación. Se usa el análisis de datos de encuestas, visitas a terreno y evaluaciones sociales para generar una amplia gama de resultados, lo que significa un valioso aporte para la eficacia de los proyectos y una identificación precisa de las áreas en que se requieren reformas.

Oportunidad de los resultados. Muchos de los componentes de evaluación se diseñaron explícitamente para el ciclo de los proyectos, fueron sincronizados para generar resultados durante las etapas de preparación de los mismos, de modo que los resultados se pudieran usar eficazmente como antecedentes para las políticas. Varios de los componentes generan datos permanentemente, en un proceso ininterrumpido de supervisión de los proyectos.

VIII. Fuentes

Jalan, Jyotsna, y Martin Ravallion. 1999. *“Income Gains from Workfare and their Distribution”*. Banco Mundial. Washington, DC. Procesado.

Ravallion, Martin. 1999. *“Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved”*. Banco Mundial. Washington, DC. Procesado.

Anexo 1.2: ¿Las microfinanzas ayudan realmente a los pobres? Nuevas evidencias de programas representativos Bangladesh

I. Introducción

Descripción del proyecto. El programa de microfinanzas para el Banco Grameen, el Comité de Adelanto Rural de Bangladesh y el Consejo de Desarrollo Rural de Bangladesh son programas de punta en relación con aquellos que se han implementado en muchos otros países. En el marco de estos programas, se otorgan pequeños préstamos a los hogares pobres que poseen menos de medio acre de tierra. Los préstamos vienen acompañados de innovadores contratos y programas de reembolso. Los programas han servido a más de 4 millones de clientes pobres en Bangladesh y aparentemente han tenido bastante éxito. Así, por ejemplo, el cuartil superior de los prestatarios del Banco Grameen consume un 15% más, tiene el doble de hijos varones en la escuela y una proporción significativamente mayor de hijas estudiando que el cuartil más bajo.

Puntos destacados de la evaluación. En la evaluación se investiga el impacto de los programas en 1.800 hogares de Bangladesh, los que entonces se comparan con un grupo de hogares de control en áreas que no cuentan con ningún tipo de microfinanciamiento. La principal contribución del estudio es demostrar que en las estimaciones simples se puede sobredimensionar significativamente la repercusión de los programas: al corregir por el sesgo de selección, se anulan ganancias aparentemente impresionantes. En la evaluación se muestra que gran parte de las ganancias percibidas resultan de las diferencias en las personas que obtienen los préstamos: suelen ser más prósperas y trabajar más que los grupos de control. Tan pronto se usan técnicas adecuadas, los préstamos no repercuten en el consumo y los niños de las áreas de los programas de hecho muestran peores resultados que aquellos de la áreas de control. El factor determinante clave es que los préstamos asignados en el marco de los programas no siguieron las directrices de calificación. De hecho, muchos de los prestatarios tenían más tierra que el límite máximo de medio acre.

En la evaluación se usa una interesante técnica de encuesta y las técnicas econométricas se aprovechan de manera imaginativa. Otro ángulo interesante es que en la evaluación se examina el efecto del impacto en la varianza y también el resultado medio, determinando que la principal ganancia de los programas es que reduce los riesgos más que aumentar los resultados medios.

II. Preguntas de la investigación y diseño de la evaluación

Los investigadores están interesados en identificar el impacto de los programas de microfinanciamiento en el logaritmo del consumo per cápita, la varianza del logaritmo del consumo, logaritmo del trabajo por adulto en los meses previos, varianza del logaritmo de trabajo por adulto, horas de trabajo de hombres adultos durante el mes pasado, horas de trabajo de mujeres adultas durante el mes pasado, porcentaje de niños

matriculados en la escuela (5 a 17 años) y porcentaje de niñas matriculadas en la escuela (5 a 17 años)

La evaluación se basa en encuestas y cubre 87 aldeas entrevistadas tres veces durante 1991 y 1992. Las aldeas se seleccionaron aleatoriamente de 5 subdistritos que sirvieron de control y 24 subdistritos en que se implementaron los programas, sobre la base de un censo y listas administrativas. En cada aldea se entrevistaron veinte hogares.

Esto permitió que los investigadores dividieran a los hogares en cinco tipos diferentes, dependiendo del criterio de calificación de tener medio acre de tierra. Es útil reproducir el esquema, ya que muestra cómo crear las variables ficticias que caracterizan la tipología y cómo abordar el sesgo de selección.

Aldea 1: Con programa			Aldea 2: de control
A No califica [b=1;e=0;c=0]		Hogares con más de ½ acre	B No calificaría [b=0;e=0;c=0]
C Califica, pero no participa [b=1;e=1;c=0]	D Participantes [b=1;e=1;c=1]		E Calificaría [b=0;e=1;c=0]
		Hogares con ½ acre y menos	

La comparación de los resultados del grupo D con los resultados del grupo C conlleva grandes problemas de selección: las cifras indican que los hogares del grupo C no participan, porque temen no ser capaces de reembolsar el préstamo. Sin embargo, si la tenencia de tierras es exógena, los grupos C y D se pueden comparar con el grupo E, porque las diferencias en los resultados dependen de la asignación de los programas, más que de la autoselección. Esto lógicamente no es así si hay diferencias entre las aldeas. Si éstas existen (debido, posiblemente, a una asignación no aleatoria), sería mejor adoptar el enfoque de diferencia en las diferencias. Así, un evaluador puede calcular los resultados medios para C y D, los resultados medios para A y luego calcular la diferencia. De igual modo, se puede calcular la diferencia entre los resultados medios para E y los resultados medios para B y entonces comparar las diferencias dentro de las aldeas.

III. Datos

Los investigadores recopilaron información sobre 1.798 hogares, de los cuales 1.538 cumplían con los requisitos para participar y 905 efectivamente participaron. Las encuestas se realizaron en 1991 y 1992 después de las cosechas de las tres principales temporadas de arroz. Las variables de interés clave fueron el consumo per cápita durante la semana anterior, el monto del crédito recibido, la cantidad de tierra que poseen, la oferta de mano de obra durante el mes pasado y las características demográficas. También se usó una fuente secundaria de información sobre transacciones de tierra, para verificar la actividad de mercado en tierras.

IV. Técnicas econométricas

Las técnicas utilizadas tenían tres componentes interesantes. El primero de ellos es la utilización de datos administrativos para verificar los supuestos clave necesarios para usar una estrategia de diseño de discontinuidad de la regresión: la exogeneidad de la tenencia de tierras. El segundo componente es un uso muy exacto de técnicas de representación gráfica no paramétrica para describir la probabilidad de cumplir con los requisitos y la probabilidad de obtener un préstamo como una función de la tenencia de tierra. Esto se combina con un análisis muy acertado de cuándo es adecuado usar un diseño de discontinuidad de la regresión, ya que el análisis gráfico sugiere que no hay un punto de equilibrio claro a 0,5 acre. Finalmente, en el estudio se usan básicamente técnicas de diferencia en las diferencias.

V. Ejecutores

Los datos fueron recopilados por el Instituto de Estudios de Desarrollo de Bangladesh por cuenta del Banco Mundial. El análisis fue realizado por el investigador Jonathan Morduch.

VI. Resultados

Los resultados parecen indicar que prácticamente todas las aparentes ganancias del programa se originaron del sesgo de selección causado por la deficiente focalización de los préstamos. Específicamente, los autores determinaron que 20% a 30% de los prestatarios tenían más del medio acre máximo que se exigía en el programa como condición para calificar, lo que sugiere la probable inclinación de los funcionarios del programa a torcer las reglas de maneras imposibles de observar. Cuando las comparaciones se restringen a los prestatarios que cumplen con la restricción en materia de tierra, con los métodos de diferencia y diferencia en las diferencias los autores encontraron que el consumo promedio en las aldeas con acceso a microfinanciamiento es inferior a la de los controles. Esto apunta a una significativa falta de focalización de los fondos del programa y por este motivo, no se pueden usar métodos de discontinuidad de regresión para analizar los impactos del programa.

La evaluación también es de utilidad para comparar resultados de diferentes técnicas econométricas: los resultados difieren marcadamente cuando se usan métodos de impactos fijos y diferencia en las diferencias o de diferencias simples. La evaluación aporta argumentos convincentes respecto a que el primero es menos adecuado cuando se utilizan diferencias de grupos beneficiarios imposibles de observar al tomar la decisión de asignación. Sin embargo, hay resultados contrapuestos en los dos enfoques respecto a una menor variación en el consumo y el ingreso debido a los programas, lo que apunta a la necesidad de contar con datos longitudinales. El impacto sobre la educación de hecho es perverso después de corregir por el sesgo de selección.

Cabe destacar, asimismo, que si bien este análisis arroja un impacto limitado del tratamiento en relación con el grupo de control, este último grupo de hecho quizás no sufrió de falta de acceso a financiamiento: posiblemente el financiamiento fue

suministrado por las ONG. Sin embargo, la utilidad del gasto de millones de dólares para subvencionar los programas de microfinanciamiento queda en entredicho.

VII. Lecciones aprendidas

De este estudio se pueden obtener varias lecciones muy importantes. La primera de ellas es la importancia de verificar si el programa funciona de acuerdo a lo prescrito. La segunda es tener presente la conveniencia del diseño de discontinuidad de regresión frente a las técnicas de diferencia en las diferencias o de diferencia simple. La tercera es la consideración del impacto de una intervención sobre el segundo momento de la distribución, como también sobre el primero, ya que la disminución del riesgo puede ser por sí mismo un resultado útil. Hay una lección más fundamental que no se aborda directamente, pero que claramente se deriva del estudio. Es una lección de economía política: si hay un fuerte incentivo para violar las reglas, las reglas serán violadas.

VIII. Fuentes

Morduch, Jonathan. 1998. *“Does MicroFinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh”*. Procesado, 17 de junio.

Véase también:

Khandker, Shahidur R. 1998. *Fighting Poverty with Microcredit: Experience in Bangladesh*. Nueva York: Oxford University Press para el Banco Mundial.

Anexo 1.3: Alimentos por educación en Bangladesh: Evaluación de un programa social focalizado cuando la asignación está descentralizada

I. Introducción

Descripción del proyecto. El Programa de alimentos por educación (FFE, *Food for Education*) de Bangladesh fue diseñado para aumentar la asistencia a la escuela primaria entregando arroz o trigo a los hogares seleccionados a modo de incentivo para los padres. Comenzó siendo un programa piloto, pero su tamaño e importancia han aumentado: su participación en el presupuestos de la División de Educación Primaria y Masiva aumentó del 11% en 1993-1994 al 26% en 1995-1996 y llegó a 2,2 millones de niños o al 13 por ciento de la matrícula total. El diseño es bastante interesante: el programa se focalizó jerárquicamente, ya que se entregaron alimentos por educación a todas las escuelas en algunas unidades geográficas económicamente atrasadas con bajos niveles de escolaridad. Luego se escogió a los hogares que recibirían los alimentos por agrupaciones locales dentro de las unidades geográficas, en base a criterios fijos, pero en cierto modo discrecionales (hogares sin tierra, hogares encabezados por mujeres y hogares de bajos ingresos). Los niños de estos hogares deben asistir por lo menos al 85 por ciento de las clases dentro de cada mes.

Puntos destacados de la evaluación. Esta evaluación es extremadamente útil, por cuanto ilustra lo que se puede hacer cuando el diseño de la intervención está orientado en absoluto hacia técnicas de evaluación normales y cuando dicha evaluación se debe llevar a cabo usando fuentes de datos existentes. De hecho, el enfoque en el programa EFE fue casi opuesto totalmente a una asignación completamente aleatoria: las áreas geográficas no sólo se escogieron porque tenían ciertas características, sino que los individuos dentro de ellas se seleccionaron porque necesitaban ayuda. Por lo tanto, puesto que el programa estaba dirigido a los más pobres entre los pobres, un análisis simple subestimaría su impacto.

El diseño de esta intervención origina un problema mayor en la creación de un escenario contrafactual, porque claramente la selección para el programa está determinada por la necesidad del mismo que tiene el hogar. La evaluación es un ejemplo de un enfoque innovador (que se puede generalizar fácilmente) para enfrentar el sesgo resultante por la dependencia en la descentralización del proceso de toma de decisiones. En resumen, puesto que el gobierno central asigna el gasto a través de áreas geográficas, pero son los agentes locales los que llevan a cabo las asignación en el lugar, en la evaluación se usan técnicas de variables instrumentales en base a la geografía para reducir el sesgo inherente al procedimiento de selección endógena. La aplicación del método arroja una estimación de impacto muy superior del programa FFE que con los enfoques de mínimos cuadrados ordinarios.

II. Preguntas de la investigación y diseño de la evaluación

La pregunta de la investigación es cuantificar el impacto del programa FFE en la asistencia escolar, medido como la tasa de asistencia para cada hogar. Hay poco a modo de un diseño de evaluación prospectiva: la evaluación se lleva a cabo con datos existentes, usando especialmente una encuesta domiciliaria representativa a nivel nacional y una encuesta comunitaria detallada. De hecho, la evaluación retrospectiva fue hecha diseñada para obviar la necesidad de una encuesta básica; en la evaluación simplemente se necesitaban encuestas que incluyeran características de los hogares y características geográficas específicas del área del hogar. En las siguientes secciones se entregan más detalles de cómo éstas se pueden estructurar para que se pueda inferir en forma viable el impacto de la intervención.

III. Datos

Los datos provienen de la Encuesta de Gastos Familiares de 1995-96, una encuesta representativa a nivel nacional realizada por la Oficina de Estadísticas de Bangladesh, donde se incluyen preguntas sobre la participación en el programa FFE y que cuenta con un componente de encuesta a nivel local. Los autores usan las respuestas sobre variables de características demográficas de los hogares, propiedad de la tierra, escuela y programas de 3.625 hogares rurales para identificar el impacto sobre la asistencia a la escuela. De hecho, en la encuesta de gastos familiares se mide directamente la asistencia a la escuela de cada niño: se registran tanto los días que el niño falta como aquellos en que la escuela está cerrada. La variable dependiente se construyó como la cantidad promedio de días de asistencia a la escuela para el hogar, como proporción de la cantidad de días factibles. Ambas partes de esta encuesta son cruciales. Por una parte, la información sobre el hogar ayuda a capturar el impacto de las características demográficas en la asistencia escolar. Por otro lado, la información sobre las características de la ubicación geográfica ayuda a modelar la estrategia de toma de decisiones del gobierno central y reducir el sesgo de selección que se menciona más arriba.

IV. Técnicas econométricas

En la evaluación se abordan dos problemas muy importantes que enfrentaron los investigadores de terreno. El primero es que la asignación del programa es descentralizada y, por ende, la decisión sobre la asignación está condicionada por variables que el econométrico es incapaz de observar, pero que sí son evidentes para las personas que toman la decisión. Esto significa que en la evaluación se requiere un medida que determine la asignación del programa a nivel individual, pero que no se correlacione con el término de error (y por lo tanto con los resultados del programa). El segundo problema es que sólo se dispone de una encuesta transversal para trabajar, sin una encuesta básica de los participantes, de manera que es difícil estimar sólo el impacto de la intervención.

La evaluación es extremadamente innovadora en cuanto a que usa el proceso de asignación en dos etapas como instrumento en sí. La característica clave para lograr esto es que los datos transversales incluyan tanto características del hogar como geográficas. En este caso en particular, el modelo es el siguiente:

$$W_i = \alpha IP_i + \beta' X_i + \eta Z_i + \mu_i \quad (1)$$

Aquí, W es el resultado del bienestar individual, X y X incluyen características del hogar y geográficas e IP , que es la ubicación del individuo en el programa, está correlacionado con el término de error. Claramente y de una importancia fundamental en la literatura sobre evaluaciones, las estimaciones de mínimos cuadrados de W estarán sesgadas.

En la evaluación se usan las diferencias geográficas en la asignación como instrumentos para cada asignación (ya que esto no está correlacionado con el término de error), al igual que las características del hogar. En consecuencia, esto caracteriza la relación como:

$$IP_i = \gamma GP_i + \pi' X_i + v_i \quad (2)$$

Es importante advertir aquí la importancia crucial de que Z contenga toda la información que se utiliza al tomar la decisión de asignación geográfica. En este caso, se usan los dos conjuntos de variables geográficas. Uno de esos conjuntos es bastante estándar y de hecho afecta directamente las propias decisiones de asistencia: variables de distancia a la escuela y el tipo y calidad de la escuela. El segundo conjunto dice relación con la decisión de asignación en sí y si bien es extensa, vale la pena mencionarla para efectos ilustrativos. Las variables incluyen la distribución de la tierra, intensidad de riego, calidad de los caminos, electrificación, distancia y tiempo hasta las sedes de la administración local y a la capital, distancia hasta los establecimientos de atención de salud y financieros, incidencia de desastres naturales, actitudes hacia la participación de la mujer en el mercado laboral, educación y planificación familiar, niveles promedio de escolaridad del jefe de hogar y cónyuge, religión predominante en la aldea y la cantidad de habitantes de la aldea. Estas variables se calculan a nivel de cada aldea y parecen predecir la selección bastante bien: una regresión probit sobre un total de 166 hogares resultó en un ajuste bastante adecuado (un pseudo- R^2 de 0,55). Esto señala que estas variables efectivamente capturan la asignación general.

Este conjunto de ecuaciones se puede modelar entonces usando mínimos cuadrados de tres etapas y comparar con los resultados de una regresión de mínimos cuadrados ordinarios.

V. Ejecutores

La evaluación fue ejecutada por Martin Ravallion y Quentin Wodon del Banco Mundial, como parte de un esfuerzo de colaboración a largo plazo entre la Oficina de Estadísticas de Bangladesh y la Unidad de Gestión Económica y Reducción de la Pobreza de la Región de Asia Meridional del Banco Mundial.

VI. Resultados

Hay diferencias claras entre ambos enfoques: el impacto estimado del Programa de alimentos por educación usando el enfoque de mínimos cuadrados de tres etapas resultó un 66 por ciento mayor que las estimaciones de mínimos cuadrados ordinarios sin controles geográficos y un 49 por ciento mayor que con los controles. En otras palabras, las estimaciones simples que sólo controlan por las variaciones entre los hogares (mínimos cuadrados ordinarios sin controles geográficos) *subestiman* considerablemente la efectividad del programa. Incluso si se incluyen controles geográficos para controlar aparentemente por la asignación geográfica no se eliminan los sesgos de asistencia. En términos sustantivos, la cantidad promedio de granos en el programa parece aumentar la asistencia en un 24 por ciento cuando se usó el método descrito más arriba.

Vale la pena mencionar que el factor clave para hacer que este enfoque sea válido es que existan suficientes variables para modelar la decisión de focalización y que estas variables sean muy parecidas a las que usaron los administradores. Si de todos modos quedan variables omitidas, los resultados seguirán estando sesgados.

VII. Lecciones aprendidas

Muchas evaluaciones no se pueden dar el lujo de diseñar una estrategia de recopilación de datos desde la base, ya sea porque la evaluación no formaba parte integral del proyecto desde el inicio o simplemente por razones de costo. Vale la pena estudiar esta evaluación por dos motivos. En primer lugar, permite documentar el nivel de sesgo que se puede producir si se utiliza un enfoque econométrico equivocado. Segundo, en ella se describe una manera válida en términos econométricos de estimar el impacto de la intervención sin el costo y desfase implícitos en una evaluación prospectiva.

VIII. Fuente

Ravallion y Wodon. 1998. *Evaluating a Targeted Social Program When Placement Is Decentralized*. Documento de trabajo de investigación de políticas 1945, Banco Mundial, Washington, D.C.

Anexo 1.4: Evaluación del fondo de inversión social de Bolivia

I. Introducción

Descripción del proyecto. El Fondo de Inversión Social (FIS) de Bolivia fue creado en 1991 como una institución financiera que promovía la inversión social sostenible en los sectores sociales, básicamente salud, educación y saneamiento. El objetivo que persigue la política es dirigir las inversiones a aquellas áreas históricamente desatendidas por las redes de servicio público, básicamente las comunidades pobres. En consecuencia, los fondos del FIS se asignan de acuerdo a un índice de pobreza municipal, pero dentro de las municipalidades el programa se basa en la demanda, respondiendo a los requerimientos de la comunidad según proyectos a nivel local. Las operaciones del FIS se descentralizaron aún más en 1994, fortaleciendo la función de los ministros sectoriales y los gobiernos municipales en el diseño y aprobación de los proyectos. El FIS boliviano fue la primera institución de su tipo en el mundo y ha servido de modelo para fondos similares establecidos desde entonces en otros países de América Latina, África y Asia.

Evaluación del impacto. A pesar de la generalizada implementación de fondos sociales durante la década de los noventa, han sido pocos los intentos rigurosos por evaluar su impacto en términos de reducir la pobreza. La evaluación del FIS boliviano, llevada a cabo en conjunto por el Banco Mundial y el mismo FIS, comenzó en 1991 y está actualmente en curso. En el estudio se cuenta con datos de encuestas básicas (1993) y de seguimiento (1997), que se combinan para permitir una evaluación del impacto antes y después. En él se incluyen evaluaciones separadas de proyectos de educación, salud y agua y su exclusividad radica en que en él se aplican una gama de técnicas de evaluación y se examinan los beneficios y desventajas de estas metodologías alternativas. Los resultados de la evaluación inicial son completos y actualmente se están presentando a los donantes y organismos gubernamentales para efectos de retroalimentación. Los resultados finales y los asuntos metodológicos se analizarán con mayor profundidad en conjunto con el informe del año 2000 del Fondo de Inversión Social, junto con un análisis de su eficacia en función de los costos.

II. Diseño de la evaluación

El proceso de evaluación del FIS boliviano comenzó en 1991 y está actualmente en curso. En el diseño se incluyen evaluaciones separadas de los proyectos de educación, salud y agua que valoran la eficiencia de la focalización del programa hacia los pobres, al igual que el impacto de sus inversiones en servicios sociales en los resultados comunitarios que se persiguen (por ejemplo, el mejoramiento de la matrícula escolar, las condiciones de salud y la disponibilidad de agua). Este proceso es ilustrativo de las técnicas de prácticas óptimas en la evaluación usando datos básicos para el análisis de los impactos. La evaluación también es innovadora en cuanto a que aplica dos metodologías de evaluación

alternativas —aleatorización y comparación pareada— al análisis de los proyectos de educación y contrasta los resultados obtenidos de acuerdo con cada método. Esto constituye un gran aporte, porque la aleatorización (selección al azar de beneficiarios del programa dentro de un grupo de individuos que cumplen los requisitos) se considera en general como el método más sólido en términos estadísticos, y sin embargo es la comparación pareada (usar un proceso no aleatorio para seleccionar un grupo de control que se “asemeje” lo más estrechamente posible a las características de los beneficiarios del programa) la que se usa mayormente en la práctica.

III. Técnicas de recopilación y análisis de datos

Los esfuerzos de recopilación de datos para la evaluación del FIS boliviano fueron intensos e incluyen la ejecución de una encuesta (básica) de inversión previa al FIS-II en 1993 y una encuesta de seguimiento en 1997. Las encuestas se llevaron a cabo en las instituciones que recibieron financiamiento del FIS y en los hogares y comunidades que se benefician de las inversiones. También se recopilaron datos similares de instituciones y hogares de comparación (grupo de control). En la encuesta domiciliaria se recogen datos de diferentes características, inclusive consumo, acceso a servicios básicos y la situación de educación y salud de cada miembro del hogar. Hay muestras separadas para proyectos de salud (4.155 hogares, 190 centros de salud), proyectos de educación (1.894 hogares, 156 escuelas), proyectos de agua (1.071 hogares, 18 proyectos) y proyectos de letrinas (231 hogares, 15 proyectos).

La encuesta domiciliaria consta de tres submuestras: (a) una muestra aleatoria de todos los hogares rurales de Bolivia, más la región del Chaco (una provincia); (b) una muestra de hogares que viven cerca de las escuelas en el grupo de tratamiento o de control para los proyectos de educación; (c) una muestra de los hogares que se beneficiarán de los proyectos de agua y letrinas.

Para analizar qué tan bien están los proyectos orientados hacia los pobres, el estudio usa los datos e información básicos (previo a la inversión del FIS) del lugar donde posteriormente se asignó la inversión FIS, para calcular la probabilidad de que los individuos fueran beneficiarios del FIS condicionados por su nivel de ingresos. Luego, en el estudio se combinan los datos de la encuesta básica y de seguimiento para estimar el impacto promedio del FIS en aquellas comunidades que recibieron una inversión del FIS, usando técnicas de regresión. Además del impacto promedio, se explora si es posible identificar las características de las comunidades, escuelas o centros de salud asociadas a impactos considerablemente mayores que el promedio.

En educación, área en la cual se asignaron aleatoriamente las inversiones FIS dentro de un grupo más grande de comunidades igualmente elegibles, se aplica el diseño de experimento aleatorizado “ideal” (en el cual el escenario contrafactual se puede observar directamente). En la inversión de salud y saneamiento, donde los proyectos no se asignaron al azar, se usa el método de “variable instrumental” para compensar la falta de un escenario contrafactual directo. Las variables instrumentales se correlacionan con la intervención, pero no tienen una correlación directa con el resultado.

IV. Resultados

Las inversiones FIS en educación y salud producen un claro mejoramiento en infraestructura y equipamiento. Los proyectos de educación afectan muy poco las tasas de deserción escolar, pero las calificaciones (puntuaciones) en las pruebas de rendimiento escolar entre los estudiantes de sexto grado son considerablemente superiores en las escuelas del FIS. En salud, las inversiones del FIS hacen aumentar las tasas de utilización del servicio de salud y reducen la mortalidad. Los proyectos de agua del FIS están asociados con leves mejorías en la calidad del agua, pero sí mejoran el acceso y la cantidad de agua y también reducen las tasas de mortalidad.

Una comparación de los resultados aleatorizados contra los de la comparación pareada en educación demuestra que el segundo enfoque produce menos grupos de comparación y tratamiento comparables y, por lo tanto, resultados menos sólidos en cuanto a discernir el impacto del programa. Como ilustración de esta conclusión, existe evidencia de mejoramientos en la infraestructura escolar (que uno claramente esperaría encontrar en las escuelas beneficiadas por el FIS) en el diseño de la evaluación aleatoria, pero no así en el diseño de la comparación pareada.

Por último, los resultados muestran que las inversiones del FIS II generalmente no están bien orientadas hacia los pobres. Los proyectos de salud y saneamiento benefician a hogares que están relativamente bien en términos de ingreso per cápita y no hay una relación entre el ingreso per cápita y los beneficios en educación del FIS.

V. Aplicación de políticas

Los resultados sobre la focalización revelan un conflicto inherente entre las metas de orientación hacia los pobres y la naturaleza del FIS basada en la demanda. Con la promulgación de la ley de participación popular en 1994, los subproyectos debían ser presentados a través de los gobiernos municipales. Los resultados de la focalización indican que incluso en un sistema muy descentralizado, es importante supervisar los procesos de focalización. En el caso boliviano, parece ser que las comunidades que tienen más probabilidades de obtener inversiones FIS son aquellas que están en mejores condiciones y son más organizadas, en lugar de las más pobres. Particularmente en el caso de los proyectos de saneamiento del FIS, el sesgo contra las comunidades más pobres puede ser difícil de corregir. La inversión en sanidad básica es más eficiente en áreas pobladas que ya tienen acceso a un sistema de agua, de modo que el proyecto puede aprovechar las economías de escala.

El hecho de que las inversiones FIS no tengan un impacto notorio en la asistencia escolar ha motivado la reestructuración de las intervenciones del FIS en este sector. En lugar de centrarse exclusivamente en proporcionar infraestructura, en los proyectos se combinarán componentes diseñados para mejorar la calidad de las escuelas. De igual modo, los desalentadores resultados en relación a la calidad del agua (lo que demuestra el nulo mejoramiento producto de los proyectos del FIS en comparación con la fuente existente

anteriormente) han generado mucha atención y se está replanteando el diseño de los proyectos en esta área.

V. Lecciones aprendidas

Eficiencia del la técnica de aleatorización. El diseño de la investigación aleatoria, en el cual se selecciona un grupo de control al azar entre posibles beneficiarios del programa, es lejos más eficiente en detectar el impacto de un programa que el método de comparación pareada para generar un grupo de control. La aleatorización se debe incorporar al diseño del programa desde su inicio en la determinación del proceso a través del cual se seleccionarán los beneficiarios del programa y la sección al azar no siempre es factible. Sin embargo, cuando los fondos del programa no son suficientes para cubrir a todos los beneficiarios, tiene sentido recurrir a la selección aleatoria de entre un grupo más numeroso de beneficiarios que cumplen los requisitos.

Importancia de institucionalizar el proceso de evaluación. Las evaluaciones pueden ser extremadamente complejas y demandar mucho tiempo. En un intento por captar rigurosamente el impacto del proyecto, la evaluación de Bolivia se realizó en el transcurso de siete años, y se lograron importantes resultados a este respecto. Sin embargo, la evaluación fue difícil de manejar por lo prolongado del tiempo y la cantidad de los diferentes actores involucrados (organismos gubernamentales e instituciones de financiamiento). La administración e implementación de una evaluación se puede agilizar incorporando estos procesos al curso normal de las actividades ministeriales locales desde un comienzo. Más aún, los enormes esfuerzos implicados se pueden restringir sólo a unos pocos programas (por ejemplo, programas grandes en los que hay enorme incertidumbre respecto a los resultados), en los cuales las compensaciones por la gestión evaluadora serían mayores.

VII. Costos y administración de la evaluación

Costos. El costo total estimado de la evaluación del FIS boliviano hasta la fecha es de \$878.000, lo que representa en 0,5% del costo total del proyecto. La recopilación de datos representa una proporción relativamente alta de este costo (69%) y el resto son gastos de viajes, tiempo del personal del Banco Mundial y consultores.

Administración. La evaluación fue diseñada por el personal del Banco Mundial y financiada en conjunto por el Banco Mundial, la agencia KfW y los gobiernos de Holanda, Suecia y Dinamarca. El trabajo de encuestas fue realizado por el Instituto Nacional Boliviano de Estadísticas y administrado por las contrapartes del FIS en la primera parte y posteriormente por el Ministerio de Finanzas en la segunda ronda.

VIII. Fuente

Pradhan, Menno, Laura Rawlings y Geert Ridder. 1998. "The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation." *World Bank Economic Review* 12 (3): 457-82.

Anexo 1.5: El impacto de los programas activos de trabajo: República Checa

I. Introducción

Descripción del proyecto. Muchos países en desarrollo se enfrentan con el problema de reconvertir o readiestrar a los trabajadores cuando se reduce el tamaño de las empresas estatales. Esto es especialmente complicado en economías en transición que también se caracterizan por tener un elevado desempleo y un estancamiento o disminución de los salarios. Sin embargo, los programas de readiestramiento no son iguales en su concepción. Algunos son simplemente indemnizaciones por despido disfrazadas para los trabajadores desplazados, otros son programas de desempleo disfrazados. Debido a ello, hay argumentos muy convincentes en favor de la evaluación de estos programas.

No obstante, los programas de capacitación son especialmente difíciles de evaluar y la evaluación checa no constituye una excepción. Lo habitual es que se instauren diferentes programas para los diferentes grupos de trabajadores. También hay numerosas formas para medir los resultados, por ejemplo en términos de empleo, trabajo por cuenta propia, ingresos mensuales e ingresos por hora. Más que en otros tipos de evaluaciones, la magnitud del impacto depende bastante del tiempo: los resultados obtenidos pueden diferir mucho dependiendo de si el período de evaluación es un mes, seis meses, un año o cinco años después de la intervención.

Puntos destacados de la evaluación. Con la ayuda de métodos de diseño casi experimentales, en esta evaluación se cuantificó el impacto de cuatro programas de mercado laboral activo (ALP, en su sigla en inglés) en la República Checa, en que se compararon participantes en estos programas con un grupo similar de no participantes. Se utilizaron tanto datos administrativos como de una encuesta de seguimiento en una evaluación ex-post de una variedad de diferentes resultados: duración del desempleo, probabilidad de empleo, trabajo por cuenta propia e ingresos. Con la ayuda de un análisis de regresión se estimó el impacto de cada uno de los cinco programas en estos resultados, controlando por las características demográficas básicas iniciales.

De esta evaluación se pudieron aprender varias lecciones importantes. Un conjunto de estas lecciones es de naturaleza práctica: cómo diseñar un evaluación altamente compleja, cómo usar los datos administrativos, cómo abordar los problemas asociados con la administración de la encuesta y los mecanismos para crear la muestra pareada. El segundo conjunto es la forma de estructurar un análisis para obtener información pertinente a las políticas a través de una evaluación detallada del impacto por subgrupo. Esto resultó en la recomendación de política de orientar los programas de mercado laboral activo hacia tipos específicos de clientes y en la conclusión de que un solo tipo de estos programas no es en absoluto eficaz en generar cambios en el empleo o los ingresos.

II. Preguntas de la investigación y diseño de la evaluación

Esta evaluación forma parte de una evaluación más amplia de cuatro países: la República Checa, Polonia, Hungría y Turquía. Todos comparten el contexto de un elevado desempleo (en parte debido a la disminución de tamaño de las empresas estatales) que se había abordado con programas pasivos de apoyo a los ingresos (como asistencia social y subsidios de desempleo). Esto se combinó con programas del mercado laboral activo, los que son el tema de esta evaluación. Los cinco ALP son Empleos con fines sociales (creación de nuevos empleos), Empleos de utilidad pública (empleos públicos de corto plazo), Programas para egresados de la escuela (subsidios para la contratación de recién egresados), Readiestramiento (capacitación laboral específica que dura unas pocas semanas hasta varios meses) y Programas para discapacitados y desfavorecidos. Este último programa es reducido y no se incluye en la evaluación.

Hay dos aspectos que hay que investigar. Uno es examinar si los participantes en los diferentes ALP tienen más facilidades para reincorporarse al mercado laboral que los no participantes y si esto varía de acuerdo a las condiciones del mercado laboral y entre los subgrupos. El segundo es determinar la eficacia en función de los costos de cada ALP y proponer mejoras.

La evaluación es de un diseño ex-post casi experimental, básicamente de grupo comparativo. El grupo de participantes se compara con un grupo calculado de no participantes (con información que se extrae de registros administrativos) de personas que se inscribieron en el Servicio de empleo, pero no fueron seleccionadas para el ALP. La idea básica es que se selecciona aleatoriamente a un individuo del grupo de participantes en el ALP. Luego se comparan los resultados de este individuo con los resultados de individuos del grupo de no participantes (sobre la base de edad, género, educación, número de meses sin empleo, tamaño de la ciudad, estado civil y tipo del último empleo). La evaluación es especialmente sólida en su análisis detallado del grupo comparativo frente al grupo de participantes.

Es inevitable que este método plantee algunos problemas, los que fueron abordados en profundidad en otros estudios (Burtless, 1995) y Heckman y Smith, 1995). Una inquietud obvia que es endémica a cualquier ensayo no aleatorio es que se pueden haber seleccionado los mejores participantes en el programa de capacitación sobre la base de características imposibles de observar o de medir por los investigadores. La segunda importante inquietud es que los no participantes pueden haber reemplazado la capacitación pública por otros tipos de capacitación en el caso del programa de readiestramiento. La tercera inquietud es que los subsidios al empleo de trabajadores puede haber causado la sustitución de un grupo de trabajadores por otro y nada más.

III. Datos

Un componente muy interesante de esta evaluación fue el uso de datos administrativos fiscales para crear el marco muestral de la encuesta. Por este motivo, el equipo visitó el Ministerio de Trabajo y Asuntos Sociales (MOLSA) en Praga y tres oficinas del mercado de empleo locales para entender tanto la administración e implementación de los ALP,

como la información administrativa sobre los participantes en estos programas. En base a esto se seleccionaron 20 distritos para la encuesta, fundamentados en la dispersión geográfica y variedad de sus características industriales. En estos distritos también existía una amplia gama de tasas de desempleo. En la encuesta se incluyeron preguntas cuantitativas sobre los resultados clave de los programas y preguntas cualitativas sobre la calificación que los participantes otorgaban al programa.

Otro valioso componente fue la implementación de una encuesta piloto en cuatro distritos. Este método de evaluación, que siempre es importante, no sólo permitió identificar los problemas técnicos, sino también un problema legal que puede presentarse a menudo con la utilización de archivos administrativos. Es el problema de la interpretación de la ley de privacidad. En este caso, el Ministerio de Trabajo y Asuntos Sociales no autorizó el envío directo por correo, sino exigió que los eventuales declarantes autorizaran a la oficina de empleo la entrega de sus direcciones. Esto retrasó el calendario de la evaluación, aumentó los costos y disminuyó drásticamente la tasa de respuestas.

La encuesta se realizó a principios de 1997 sobre una muestra aleatoria de 24.973 inscritos en la oficina de empleo que fueron contactados. De éstos, 9.477 participaron en el ALP en el período 1994-95. La tasa de respuestas de los no participantes fue un 14% y de los participantes, un 24,7%, lo que arrojó un número total de 4.537 declarantes. La baja tasa de respuestas se puede atribuir directamente a la normativa legal: la mayoría de las personas no respondieron a la solicitud inicial. Sin embargo, la tasa de respuestas fue muy elevada entre los que autorizaron la entrega de su dirección. Lo peor es que desconoce el sesgo que esto pueda haber causado.

IV. Técnicas econométricas

La dificultad inherente a la medición de la naturaleza temporal y la complejidad de los resultados del mercado laboral se ve reflejada en el uso de siete medidas diferentes de los resultados: porcentaje de los que actualmente están empleados, porcentaje de los que actualmente trabajan por cuenta propia, porcentaje de los que trabajaron alguna vez, duración del desempleo, duración de los subsidios de desempleo, total de subsidios de desempleo y ingresos mensuales actuales.

Sin embargo, el método de evaluación fue bastante directo en su utilización de diferencias simples entre grupos, por una parte, y mínimos cuadrados ordinarios con variables ficticias específicas por grupo, por la otra, para medir el impacto de las intervenciones. Se calculó el impacto general, seguido de los impactos estimados para las categorías de cada uno de los subgrupos (edad, género, educación y, para los resultados en materia de ingresos, tamaño de la empresa). Este último análisis fue especialmente útil, ya que identificó a subgrupos de individuos que efectivamente se vieron afectados de manera diferente por las intervenciones, llevando a implicaciones de política muy diferentes. De hecho, una recomendación muy importante que surgió de la evaluación fue que los ALP se orientaran más rigurosamente hacia grupos beneficiarios específicos.

V. Ejecutores

La evaluación formó parte de una evaluación comparada de programas activos de trabajo entre cuatro países, destinada expresamente a entender el impacto de los ALP en el contexto de diferentes condiciones económicas. Fue supervisada por un comité consultivo de proyectos que contó con representantes del Banco Mundial, de cada uno de los cuatro países, de las agencias de financiamiento externas y de los contratistas en asistencia técnica (Abt Associates y el Upjohn Institute).

El equipo contrató a una empresa encuestadora privada para realizar la encuesta misma, tanto por motivos de calidad como para disminuir una posible intimidación en caso de que fuera la oficina de trabajo local la que efectuara la encuesta. Cabe insistir en que la credibilidad del estudio podría haberse visto comprometida si el servicio de trabajo hubiera estado a cargo de ejecutar la encuesta. En efecto, este problema de riesgo moral suele ser importante si el organismo a cargo de la capacitación es simultáneamente responsable de recopilar información sobre los resultados de la misma.

VI. Resultados

Los resultados son los típicos para las evaluaciones de programas de capacitación. Algunas intervenciones parecen haber influido de alguna manera (aunque relativamente débil) para algunos tipos de trabajadores y en algunas situaciones. Un aspecto positivo de la evaluación es que al menos se identificó un programa en el que parece haberse desperdiciado el dinero: no mostró efecto alguno a nivel global ni en algún subgrupo. Otro punto sólido a favor de la evaluación es su presentación, la que es especialmente importante si ésta llegará a las manos de las autoridades normativas. En ella se incluyen cuadros para cada programa que resumen los beneficios combinados en materia de salarios y empleo, tanto a nivel agregado como para cada subgrupo.

Un aspecto muy negativo es que no se realizaron análisis de costos-beneficios, a pesar de las promesas iniciales. Estos análisis habrían sido extremadamente útiles para comparar la información sumaria de los beneficios con el costo combinado explícito e implícito del programa. Así, si bien los evaluadores determinaron que un programa aumentó la probabilidad de empleo en todas las categorías, es necesario mencionar que fue a cambio de un programa de capacitación de nueve meses. En un cálculo completo de la tasa de rentabilidad de la inversión, se habría combinado el costo explícito del programa con el costo de oportunidad del tiempo de los participantes y esto se habría comparado con el aumento de ingresos y empleo.

VII. Lecciones aprendidas

De este estudio se pueden aprender varias lecciones importantes. Algunas de las principales son los componentes pragmáticos que se analizaron en la introducción, especialmente la importancia de tener en cuenta el entorno político al momento de diseñar un sistema de evaluación. La imposibilidad de convencer al Servicio de Empleo sobre la importancia del proyecto de evaluación significó que el instrumento de estudio se vio seriamente comprometido. Segundo, el estudio entrega una demostración útil de la construcción de una muestra pareada. Finalmente, la evaluación es una buena ilustración de la importancia de realizar un análisis no sólo a nivel agregado, sino también de subgrupos, con la consiguiente posibilidad de realizar fructuosas intervenciones orientadas hacia grupos beneficiarios específicos.

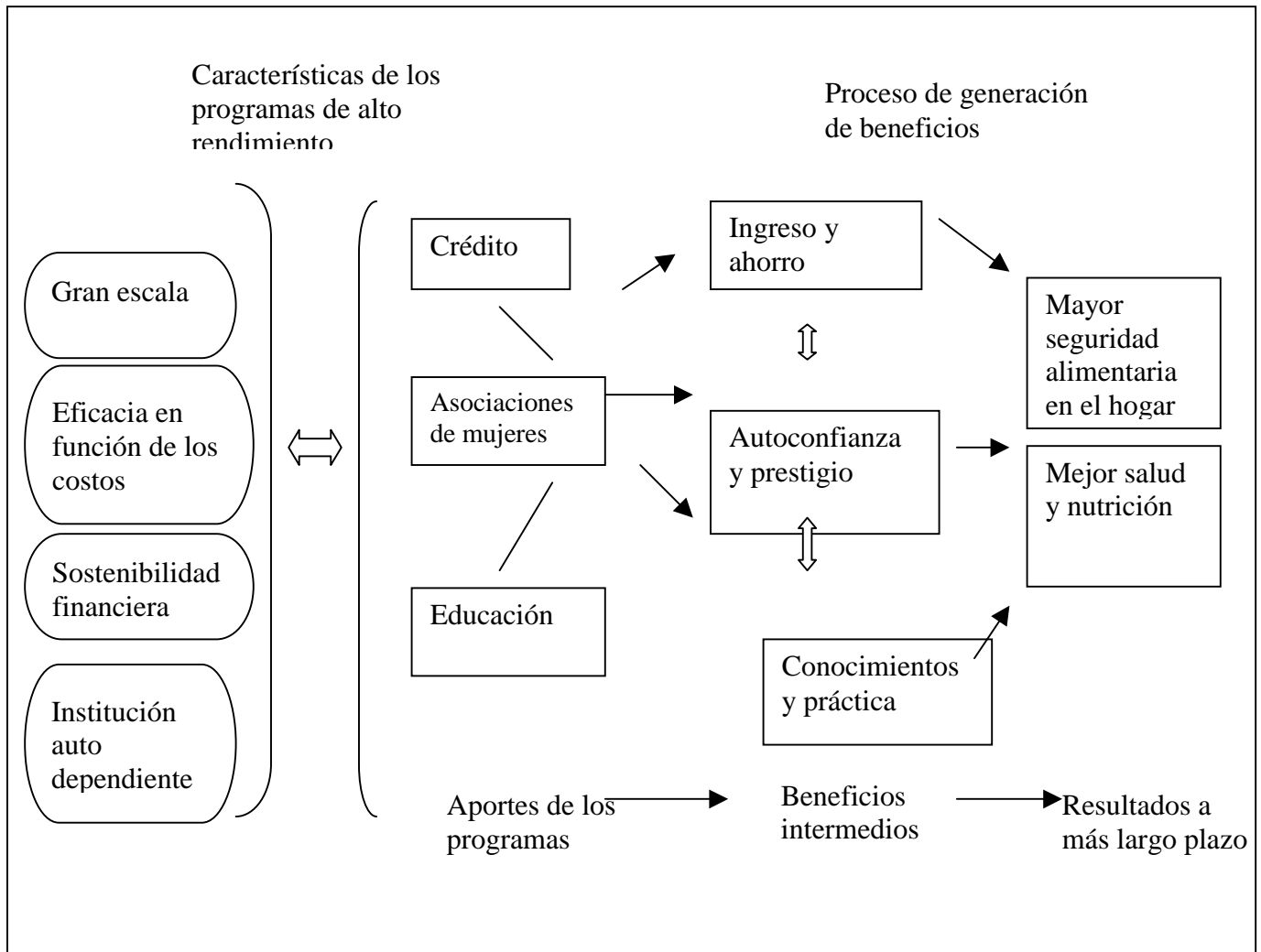
VIII. Fuentes

Benus, Jacob, Grover Neelima, Jiri Berkovsky y Jan Rehak. 1998. *Czech Republic: Impact of Active Labor Market Programs*. Cambridge, Mass., y Bethesda, MD.: Abt Associates, mayo

Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research". *Journal of Economic Perspectives* 9(2):63-84.

Heckman, James J., y Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments". *Journal of Economic Perspectives* 9(2) : 85-110.

Esquema usado para diseñar la evaluación de los programas activos de trabajo checos



Anexo 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

I. Introducción

Descripción del proyecto. En el programa Crédito con Educación se combinan elementos del programa del Banco Grameen Bank con educación sobre nociones básicas de salud, nutrición, oportunidad y espaciamiento de los nacimientos y capacidad de manejo de pequeñas empresas. Su objetivo es mejorar el estado nutricional y seguridad alimentaria de los hogares pobres de Ghana. *Freedom from Hunger*, junto con el Programa de Nutrición Internacional de la Universidad de California, Davis, proporcionó servicios de Crédito con Educación a mujeres pobres de áreas rurales en el distrito oriental de Shama Ahanta de la Región Occidental de Ghana. Se formó una asociación con cinco bancos rurales para prestar estos servicios: hasta marzo de 1997 se habían otorgado más de 9.000 préstamos por un valor total de US\$600.000, con una tasa de reembolso inferior al 92%.

Puntos destacados de la evaluación. Esta evaluación es interesante por tres motivos. En primer lugar, el diseño de la muestra fue bastante adecuado, ya que el programa se aplicó en 19 comunidades y se recopilaron datos de tres grupos de mujeres muestrales diferentes: aquellas que participaron al menos durante un año, las que no participaron, pero pertenecían a comunidades donde se implementó el programa y aquellas pertenecientes a comunidades de control. En segundo lugar, en el estudio había una clara descripción de su enfoque implícito: se identificaban y evaluaban resultados intermedios y a un plazo más largo. Finalmente, presentaba una adecuada combinación de resultados cualitativos y cuantitativos, donde a menudo se destacaban los resultados cuantitativos con un ejemplo ilustrativo anecdótico.

II. Preguntas de la investigación y diseño de la evaluación

Las preguntas de la evaluación se centraron en los impactos del programa sobre a) el estado nutricional de los niños, b) la capacidad económica de las mujeres (ingreso, ahorros, tiempo) para invertir en la alimentación y cuidados de salud, c) los conocimientos y adopción de prácticas de amantamiento, destete y manejo y prevención de diarrea por parte de las mujeres y d) capacidad de la mujeres de ofrecer a sus hijos una dieta sana.

Para este fin, en la evaluación se separaron las metas finales de una mayor seguridad alimentaria en el hogar y un mejor estado nutricional, de los beneficios intermedios de cambiar los hábitos, reducir la pobreza y habilitar u otorgar mayor poder a las mujeres.

Se utilizó un diseño casi experimental para aplicar en terreno dos encuestas (en 1993 y 1996) que permitieran evaluar el impacto de la estrategia en el estado nutricional de los

niños, la capacidad económica de las madres, el empoderamiento de las mujeres y la adopción de prácticas de salud y nutrición infantil por parte de las madres. En el primer período se entrevistó a un total de 299 pares de madres y sus hijos y 290 pares diferentes en el segundo período, recopilando tanto información cualitativa como cuantitativa.

El diseño de la evaluación fue bastante complejo. El banco rural The Lower Pra identificó a 19 comunidades que todavía no habían recibido servicios de Crédito con Educación y los consultores dividieron a las comunidades en grandes y pequeñas (más o menos de 800 habitantes) y luego nuevamente de acuerdo a su ubicación en la cercanía de una carretera principal o no. Dentro de cada estratificación, 13 de las 19 comunidades se asignaron a un grupo de tratamiento o de control. A tres se les dio el tratamiento por motivos políticos y se seleccionaron tres comunidades como controles comparativos de las tres seleccionadas políticamente sobre la base de su proximidad, desarrollo comercial, tamaño y acceso a carreteras principales. Se descartaron dos comunidades debido a la falta de interés que presentaban y el reducido número de comunidades en la clasificación. Por consiguiente, en el estudio de seguimiento sólo se entrevistó a 14 comunidades.

Para las encuestas básicas se seleccionaron diez pares de madres y sus hijos, con niños entre 12 y 23 meses de edad, de comunidades pequeñas y treinta pares de comunidades grandes. Esto planteó dos problemas importantes. En primer lugar, esta construcción no permitió el seguimiento de las mismas mujeres a través del tiempo, porque pocas mujeres de la encuesta básica también tenían niños de esa edad en la encuesta de 1996. En segundo lugar, la restricción de edad limitó de tal forma la segunda muestra que tuvo que ampliarse a mujeres con niños menores de tres años en 1996. Este complejo diseño de la evaluación tuvo la importante ventaja de permitir la clasificación de las mujeres en la muestra básica como futuras participantes y futuras no participantes.

Se entrevistó a tres tipos de mujeres: participantes, no participantes en las comunidades del programa y residentes en las comunidades de control. Se incluyó a todas las participantes, mientras que los últimos dos tipos se seleccionaron aleatoriamente mujeres con niños menores de tres años. Cabe señalar que el tamaño de la muestra completa (360) se calculó sobre la base de las desviaciones estándares que se encontraron en estudios previos, un requisito de que la muestra permitiera detectar una diferencia de 0,4 en los valores de la puntuación z de los grupos de control y grupos beneficiarios y con un nivel de significancia previsto de 0,05 y una potencia de 0,8.

III. Datos

Se recopilaron datos cuantitativos y cualitativos sobre el hogar, la madre y el hijo, centrándose tanto en medidas intermedias y a largo plazo y especialmente en la naturaleza multidimensional de los resultados.

Para los resultados intermedios, esto llevó a un conjunto de preguntas que intentan medir la capacidad económica de las mujeres (ingresos, utilidades, contribución al ingreso del hogar, ahorros, capacidad empresarial y gastos en alimentos y hogares). De manera similar, con otro conjunto de medidas se abordaban los conocimientos de las mujeres sobre salud y nutrición (amamantamiento, alimentación infantil, tratamiento y prevención

de diarreas e inmunizaciones). Con otro conjunto más se trató de captar la habilitación de las mujeres (autoconfianza y esperanzas en el futuro, posición y capacidad de toma de decisiones en el hogar y prestigio y redes sociales en la comunidad). Para los resultados finales como estado nutricional y seguridad alimentaria se utilizaron medidas más directas (medidas antropométricas para el primero, preguntas sobre hambre para el segundo).

Si bien se programó un tamaño de muestra de 360 pares de madres y sus hijos, en la primera encuesta sólo se entrevistaron 299 (principalmente porque se descartaron dos comunidades) y 290 en la segunda. Se compararon las características de las madres y los hogares de cada uno de los tres grupos y no se encontraron diferencias significativas.

IV. Técnicas econométricas

Las técnicas econométricas que se usaron son bastante directas y exploraron la solidez del diseño de la encuesta. Se calculó la media del grupo para cada una de las variadas medidas de resultado que se utilizaron y luego se realizaron pruebas t para examinar si las diferencias entre los controles y los participantes eran significativas. En su esencia, se trata de un método de evaluación de diferencia simple. Esto está adecuadamente complementado con gráficos.

Sin embargo, quedaron una serie de preguntas importantes sin abordar. En primer lugar, el diseño de la muestra fue por conglomerados y, dado que los resultados de cada par individual de madre e hijos casi por construcción estarán correlacionado con los demás en la comunidad, los errores estándares quedarán sesgados a la baja y el estadístico de t espuriamente sesgado al alza. En el caso extremo, en el que todos los resultados individuales están perfectamente correlacionados entre sí, el tamaño de la muestra es en realidad 17 más que 300. Esto llevará la significancia a resultados que de hecho pueden no ser significativos. En segundo lugar, aunque el diseño explícitamente fue estratificado, no se abordó el impacto de esta estratificación: si se beneficiaron más las comunidades grandes o pequeñas o si las comunidades cercanas a una carretera estaban en mejores condiciones que aquellas muy alejadas de una carretera. Esto es especialmente sorprendente, ya que se presume que el objetivo de este diseño de la muestra era examinar sus implicaciones para las políticas. En tercer lugar, aunque se analizan los problemas de sesgo de selección, no se presenta ningún análisis formal o corrección para este problema esencial. Finalmente, aunque habían diferencias significativas en las tasas de no respuesta de los ítems, lo que apunta a la posibilidad de un sesgo de selección incluso dentro de la encuesta, esto no se abordó ni se analizó.

V. Ejecutores

El programa Crédito con Educación fue desarrollado por un instituto internacional sin fines de lucro, *Freedom from Hunger*, el que colaboró con el Programa de Nutrición Internacional de la Universidad de California, Davis, para evaluarlo. El instituto colaboró con el banco rural The Lower Pra (un banco autónomo fiscalizado por el Banco de Ghana) y posteriormente con otros cuatro bancos rurales de Ghana para suministrar el

programa. El banco rural The Lower Pra participó en la identificación y selección de las comunidades que se entrevistarían.

VI. Resultados

En general se alcanzaron los objetivos intermedios: si bien no hubo un aumento en los ingresos y gastos de las mujeres, sus capacidades empresariales y de ahorro fueron significativamente superiores. También habían mejorado en general los conocimientos de las mujeres sobre salud y nutrición y era más probable que se sintieran con poder. En términos de las metas finales, la evaluación parece indicar que el programa mejoró la seguridad alimentaria y el estado nutricional de los niños en los hogares, pero no el estado nutricional de las madres.

VII. Lecciones aprendidas

Una contribución clave de la evaluación es el diseño muy interesante de la muestra: la estratificación y la selección de los grupos de participantes y no participantes con respecto a su futura participación es un método de evaluación muy útil. Otra lección es el uso productivo de muchas dimensiones de los resultados, en ocasiones factores en alto grado no cuantitativos, como el empoderamiento de las mujeres. La otra lección clave es el valor de los datos no cuantitativos para ilustrar la validez de las inferencias cuantitativas.

VIII. Fuente

MkNelly, Barbara, y Christopher Dunford (en colaboración con el Programa de Nutrición Internacional de la Universidad de California, Davis). 1998. *"Impact of Credit with Education on Mothers' and their Young Children's Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana"*. Documento de estudio *Freedom from Hunger* N° 4, marzo.

Anexo 1.7: Libros de texto y puntuaciones de pruebas: Evidencia de una evaluación prospectiva en Kenya

I. Introducción

Descripción del proyecto. Evaluar el impacto de diferentes tipos de gastos en educación en los resultados estudiantiles es particularmente importante en países en desarrollo. En los estudios anteriores parecen indicar que la entrega de libros de texto es una manera eficaz en función de los costos para aumentar las puntuaciones (calificaciones) en las pruebas y Kenya, dada la aguda falta de recursos disponibles para los educadores, constituye un buen estudio de caso. Los evaluadores observan que sólo uno de cada seis niños en tercer, cuarto y quinto grado tiene libros de texto; esta proporción aumenta a uno cada cuatro en grados superiores. Además, las instalaciones físicas son extremadamente pobres, a consecuencia de lo cual muchos niños deben sentarse en el suelo para aprender.

En la evaluación se estimó el impacto sobre el aprendizaje de un programa llevado a cabo en 1996, en el cual se entregaron libros de texto a todos los grados de un subconjunto seleccionado al azar de 25 de cada 100 escuelas primarias rurales de Kenya. Se entregaron libros de texto en inglés desde el tercer al séptimo grado, con una relación de 6 libros para 10 niños; y libros de ciencia al 60 por ciento de los estudiantes en octavo grado. Además, a cada clase se le entregó una guía para el maestro. Se tomaron pruebas de rendimiento a los estudiantes antes de distribuir los textos y nuevamente después de diez meses. También se tomaron las mismas pruebas en las escuelas de control. En este enfoque se combina un diseño aleatorio con comparaciones reflexivas.

Puntos destacados de la evaluación. Esta evaluación constituye una magnífica ilustración de cómo crear y ejecutar un buen diseño de encuesta y luego hacer un seguimiento con técnicas econométricas adecuadas. Su principal fortaleza radica en demostrar cómo inferir resultados por niveles con datos lineales, el uso de estimadores de diferencia en las diferencias, cómo abordar los sesgos de selección y de atrición, al igual que asuntos relativos a errores en la medición y desplazamiento. Otro componente muy interesante de la evaluación es su enfoque en el impacto de la intervención sobre los estudiantes en todas las partes de la distribución. Por último, el reconocimiento y análisis de los posibles efectos secundarios es un muy buen ejemplo de cómo examinar todas las dimensiones de una intervención.

II. Preguntas de la investigación y diseño de la evaluación

El enfoque principal que se sigue en la investigación es evaluar el impacto de los libros de texto en los resultados de aprendizaje. Debido a que se trata de un concepto complejo, los resultados se miden como la diferencia entre escuelas con libros de texto y escuelas de comparación, en varias dimensiones: puntuaciones postpruebas, ganancias en las puntuaciones de pruebas, diferencias entre combinaciones asignatura-grado que

recibieron y no recibieron libros de texto y actividad de estudiantes y maestros. En la evaluación también se consideraron otros efectos secundarios (que a menudo se ignoran), particularmente la posibilidad de que el otorgamiento de dicho subsidio reduciría la participación de los padres, en especial en términos de desplazar otras formas de generar recursos.

El diseño de la evaluación es bastante complejo. El Ministerio de Educación escogió 100 escuelas de bajos recursos para la intervención de 1995. Estas escuelas se dividieron en cuatro grupos; en primer lugar sobre la base de su ubicación geográfica, luego en orden alfabético dentro de su ubicación. Luego se produjo una asignación ordenada alfabéticamente de cada escuela a cada uno de los cuatro grupos. La entrega de libros de texto se llevó a cabo en etapas. El primer grupo los recibió en 1996, el segundo en 1997 y así sucesivamente. Se entregaron libros de matemáticas, inglés y ciencias a diferentes grados, fundamentalmente del tercer al séptimo grado.

III. Datos

Se tomaron exámenes de matemáticas, inglés y ciencias a los estudiantes de todos estos grados en cada una de las 100 escuelas antes de distribuir los libros de texto. Sin embargo, en la evaluación en sí se utilizan exámenes previos tomados en los grados tres al siete en octubre de 1996 y exámenes posteriores en octubre de 1997. Por lo tanto, hay datos sobre aproximadamente 8.800 estudiantes (en todos los grados) para cada asignatura en las 100 escuelas y un total de más de 26.000 observaciones. Debido a que sólo 25 escuelas recibieron los libros de texto durante este período, los estudiantes de estas escuelas se convierten en el grupo “con libro de texto”; las demás 75 constituyen el grupo de comparación. Además de las puntuaciones en las pruebas, también se recopilaban datos sobre las finanzas de la escuela y sobre los métodos pedagógicos.

La información sobre la utilización de los libros de texto en la sala de clases fue recopilada por observadores experimentados que visitaron cada escuela y tomaron notas acuciosas sobre ocho posibles actividades en el aula (desde actividades generales de los maestros y estudiantes hasta el uso de los textos de parte de los maestros y estudiantes). Estas notas abarcaron un período de 15 minutos y luego se utilizaron para sacar porcentajes del tiempo utilizado por los maestros y estudiantes en cada actividad para un total de 551 horas de clase. Personal en terreno entrevistó a cuatro o cinco estudiantes de cada clase, los que luego llenaron cuestionarios sobre la base de sus respuestas.

Por último, se recopilaban datos sobre las finanzas de las escuelas de un cuestionario de 1997 a escuelas y comités escolares, donde se preguntaba sobre las actividades para generar recursos.

IV. Técnicas econométricas

Vale la pena destacar los interesantes aspectos que generó esta técnica de muestreo. Es probable que las puntuaciones de pruebas dentro de una escuela estén correlacionadas entre sí, al igual que las puntuaciones al interior de una clase. De igual modo, se

correlacionarán los puntuaciones de pruebas en diferentes asignaturas de un mismo niño. La intervención también se puede evaluar en términos de su impacto sobre los resultados en los niveles de aprendizaje de los estudiantes o sobre las ganancias en aprendizaje de los estudiantes. En general, el efecto de una intervención debe ser sólido a diferentes técnicas econométricas y diferentes formas de observar los datos, lo que efectivamente sucedió en este caso.

La evaluación primero proporciona estimaciones de una simple regresión de nivel de variable ficticia, con variables ficticias de tratamiento para cada combinación grado-asignatura con efectos aleatorios de escuela, grado y asignatura (la variable dependiente es el cambio en las puntuaciones de las pruebas entre las pruebas previas a las posteriores). Una característica atractiva de esto es que las variables ficticias se pueden combinar de formas bastante útiles:

- Agrupando varios grados para estimar el impacto de los libros de texto para una asignatura
- Agrupando todas las puntuaciones de las pruebas para estimar el impacto promedio de los textos para un grado; y
- Agrupando todos los grados y asignaturas para estimar el impacto promedio ponderado de los libros de texto para todos los grados y asignaturas.

Ciertamente, la estructura de los efectos aleatorios varía con cada enfoque y la evaluación es muy clara en este componente.

La evaluación continúa luego con un enfoque de diferencia en las diferencias, el que es relativamente directo en cuanto a que simplemente compara las puntuaciones antes y después de las pruebas entre las escuelas de control y de tratamiento.

El tercer enfoque, que es un poco más complejo debido a que explota la variación dentro de las escuelas, merece ser analizado. La regresión aplicada aquí implica hacer una regresión de las puntuaciones de pruebas en variables ficticias que capturan si los estudiantes estaban o no (a) en una escuela con libros de texto y (b) en una combinación asignatura-grado que recibió un libro de texto. Esto reduce los problemas que se presentan debido a la heterogeneidad de las escuelas, al igual que los problemas relativos a la selección de las muestras; en este último caso, debido a que captura el efecto sobre las puntuaciones en pruebas para el mismo estudiante, dependiendo de si el estudiante recibió o no el libro. Sin embargo, sí supone que es posible sumar y restar las puntuaciones de pruebas en diferentes combinaciones grado-asignatura y este supuesto tan poderoso puede ser el motivo por el cual con este enfoque se obtienen resultados muy diferentes.

Un tema recurrente en las evaluaciones es del deseo de capturar no sólo el efecto promedio de la intervención, sino que también el efecto en los subgrupos de receptores. Esta evaluación permite visualizar de manera clara el uso de los términos de interacción y regresión de quintiles. El primer enfoque implica la interacción entre las puntuaciones de pruebas iniciales y las variables ficticias de libros de texto para capturar el efecto de los

libros de texto en mejores contra peores estudiantes, usando tanto valores reales como instrumentales (las puntuaciones de pruebas iniciales se correlacionan con el término de error, causando un sesgo). El segundo enfoque, que implica usar una regresión de quintiles, también es útil y cada vez más difundido. Más específicamente, debido a que la regresión de mínimos cuadrados sólo captura el impacto promedio del programa de libros de texto, las regresiones de quintiles permiten que el efecto del tratamiento varíe dependiendo del lugar en que se encuentra el estudiante dentro de la distribución.

Otro elemento particularmente sólido en la evaluación es el de proporcionar una aplicación de cómo buscar sesgos de selección y atrición. La principal fuente de posibles problemas en esta intervención son las diferentes tasas de promoción y repetición entre las escuelas de comparación y con libros de texto. Por ejemplo, los niños pueden ser promovidos diferencialmente del grado 2 (un grado sin libro de texto) al grado 3 (un grado con libro de texto) en escuelas con libros de texto. Una promoción diferencial produce un sesgo descendente en los resultados de aquellas clases en las que se incorporan los peores alumnos y posiblemente produce un sesgo ascendente en las clases de las cuales provienen. Estos dos efectos fueron capturados por la evaluación reestimando el modelo en dos formas: sacando a todos los repitentes de ambos conjuntos de escuelas y sacando a los peores estudiantes de cada grado. La solidez de los resultados en ambos enfoques confirmó el impacto de la intervención.

Por último, en una ilustración de cómo considerar la importancia de los efectos secundarios, en la evaluación se cuantificó el impacto de la entrega de libros de texto en la generación de fondos de parte de los padres. Se descubrió que la intervención efectivamente desplazó los aportes de los padres: el monto de la ayuda no de ICS recibida por las escuelas de comparación fue de \$465 y para las escuelas con libros de texto fue de \$267 (el valor promedio de los libros de texto de ICS es \$485). Se usaron análisis de regresión simples y también se investigó y confirmó la hipótesis de que las escuelas más pequeñas tenían más desplazamiento que las escuelas grandes.

Ejecutores. El proyecto fue financiado por una organización sin fines de lucro holandesa, International Christelijk Steunfonds. La evaluación fue realizada por un profesor del Instituto de Tecnología de Massachusetts (Kremer) y dos economistas del Banco Mundial (Paul Glewwe and Sylvie Moulin). Parte de los costos fueron cubiertos por la National Science Foundation y el comité de investigaciones del Banco Mundial.

V. Resultados

El resultado de esta evaluación entró en directa contradicción con los resultados de otras evaluaciones de intervenciones con libros de texto. El resultado fundamental fue que los libros de texto en promedio no influían considerablemente en los resultados de aprendizaje, pero sí tenían un efecto significativo en los mejores estudiantes. Dicho resultado fue sólido a diferentes técnicas de estimación y cortes de los datos.

VI. Lecciones aprendidas

La lección más útil que resulta de esta evaluación es la importancia de usar diferentes técnicas econométricas para comprobar la solidez de los resultados empíricos. Aun cuando la recopilación de datos estuvo cercana al ideal, es importante que el impacto estimado de la intervención permanezca relativamente igual con diferentes supuestos econométricos y especificaciones de modelos. La aplicación de una regresión de quintiles y términos de interacción también fue una forma muy útil de analizar el impacto en diferentes subgrupos de la población. Por último, es importante buscar e identificar efectos secundarios, en este caso, la posibilidad del desplazamiento.

VIII. Fuente

Glewwe, Paul, Mihael Kremer y Sylvie Moulin, 1998. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya". Grupo de Investigaciones sobre el Desarrollo (DECRG), Banco Mundial, Washington, D.C., Procesado.

Anexo 1.8: Evaluación del proyecto de extensión agrícola de Kenya

I. Introducción

Descripción del proyecto. El primer Proyecto Nacional de Extensión (NEP-I) de Kenya introdujo el sistema de Capacitación y Visita (T&V, *Training and Visit*) en la gestión de los servicios de extensión agrícola en 1983. El proyecto cumplía el doble objetivo de promover el desarrollo institucional y prestar servicios de extensión a los agricultores, con la meta de aumentar la productividad agrícola. En 1991 le siguió el NEP-II, apuntando a consolidar los avances logrados con el NEP-I, a través de un aumento del contacto directo con los agricultores, mejoramiento de la pertinencia de la información y tecnologías de extensión, actualización de las capacidades del personal y de los mismos agricultores y perfeccionamiento del desarrollo institucional.

Evaluación del impacto. El desempeño del sistema de extensión de Kenya ha sido controvertido y forma parte de un debate más amplio sobre la eficacia en función de los costos del enfoque T&V frente a la extensión. A pesar de la intensidad del debate, la importante función que cumplen los servicios de extensión agrícola dentro de la estrategia de desarrollo del Banco Mundial para África y el enorme caudal de inversiones realizadas, se han realizado muy pocos intentos rigurosos por medir el impacto de la extensión de T&V. En el caso de Kenya, el debate se ha incrementado debido a estimaciones de retornos muy elevados para la T&V que arrojó un estudio anterior y la falta de resultados visibles y convincentes, incluyendo el pobre desempeño de la agricultura de Kenya en los últimos años.

Las controversias (entre el Departamento de Evaluación de Operaciones y la Región de África del Banco Mundial) respecto al desempeño del NEP-I continúan en espera de los resultados de esta evaluación, en la que se utiliza un enfoque empírico riguroso para estimar el impacto del programa sobre el desempeño agrícola. Usando el marco administrativo basado en los resultados, en la evaluación se examina el impacto de los servicios del proyecto sobre la productividad y eficiencia agrícola. En ella también se desarrollan componentes para medir los resultados (es decir, conciencia de los agricultores y aceptación de nuevas técnicas) y volumen (por ejemplo, frecuencia y calidad del contacto) del programa, para evaluar el desempeño del sistema de extensión y para confirmar su impacto real o posible.

II. Diseño de la evaluación

La estrategia de esta evaluación es un ejemplo de técnicas de prácticas óptimas en la utilización de una amplia gama de métodos de evaluación para estimar la implementación, volumen e impacto del programa en la productividad y eficiencia agrícola (no hay intentos por estudiar el impacto sobre el bienestar de los hogares, lo que

probablemente se vería afectado por una cantidad de factores que trascienden el ámbito de las actividades de capacitación y visitas). La evaluación se basa en métodos tanto cuantitativos como cualitativos, de modo que los rigurosos hallazgos empíricos sobre el impacto del programa se pueden complementar con evaluaciones de los beneficiarios y entrevistas al personal donde se destacan las cuestiones prácticas en el proceso de implementación. En el estudio también se aplicó el método de valuación contingente para indagar la voluntad de los agricultores de pagar por servicios de extensión [el método de valuación contingente indaga sobre los valores de uso y no de uso de los individuos para una variedad de bienes y servicios públicos y privados; se pide a los entrevistados que declaren su voluntad de pagar (aceptar) para evitar (aceptar) un cambio hipotético en la prestación de los bienes o servicios, es decir, el resultado “contingente”. En este caso, se les preguntó a los agricultores cuánto estarían dispuestos a pagar por servicios de extensión agrícola permanentes en caso que el gobierno dejara de proporcionarlos].

La evaluación cuantitativa se complica por el hecho de que el sistema T&V se introdujo a escala nacional, impidiendo hacer una comparación con programa y sin programa (grupo de control). Por lo tanto, con la metodología de la evaluación se buscaba explotar los datos domiciliarios de la producción agrícola anteriores al proyecto para efectuar comparaciones limitadas antes y después con la ayuda de métodos de datos de panel. Con este fin, los datos disponibles sobre los hogares se complementaron con una reciente encuesta para formar un panel. No se pudieron llevar a cabo las evaluaciones de los beneficiarios diseñadas para este estudio, pero la evaluación se basa en los hallazgos pertinentes de dos recientes evaluaciones de beneficiarios realizadas en Kenya. El estudio es digno de destacar por basarse en una gama de fuentes de información previas de Kenya (encuestas domiciliarias, evaluaciones participativas, entre otras), complementadas por una gestión de recopilación de datos más amplia para los efectos de la evaluación.

III. Técnicas de recopilación y análisis de datos

El enfoque de la evaluación se basa en distintas fuentes de datos cualitativos y cuantitativos anteriores. La evaluación cuantitativa se basa en gran medida en una encuesta domiciliaria de 1998 realizada por el Departamento de Evaluación de Operaciones del Banco Mundial. A partir de esta encuesta se generan datos de panel, volviendo a visitar la mayor cantidad de hogares que se pudieron ubicar de una encuesta domiciliaria realizada en 1990 por el Departamento Técnico para África, la que a su vez se basó en una submuestra de la Encuesta de Presupuesto Familiar Rural de 1982 (estas tres encuestas crean un conjunto de datos de panel para una cantidad aproximada de 300 hogares; en las encuestas se cubren temas como la demografía del hogar, las características de la explotación agrícola y datos de insumos-productos sobre producción agrícola; en la encuesta de 1990 y 1998 también se recopiló información sobre contactos con servicios de extensión, inclusive temas de toma de conciencia y aceptación de mensajes de extensión). Estos datos se complementan con una encuesta realizada al personal de extensión, varias revisiones recientes del servicio de extensión realizadas o encargadas por el Ministerio de Agricultura y discusiones de grupos representativos particulares con el personal de extensión. El estudio también se basa en dos evaluaciones

recientes de los beneficiarios: un estudio de 1997 realizado por ActionAid Kenya, con el cual se indagó sobre las perspectivas de los usuarios y posibles usuarios de los servicios de extensión de Kenya; y una Estudio Participativo sobre Pobreza de 1994, donde se indagó sobre los servicios públicos, incluyendo la extensión, que fue realizado en conjunto por el Banco Mundial, la British Overseas Development Administration, la Fundación Médica y de Investigación de África, UNICEF y el Gobierno de Kenya.

En el análisis se evalúa tanto el proceso de implementación como el resultado del programa de capacitación y visita de Kenya. Se evalúan los procesos institucionales basándose en datos secundarios y cualitativos: encuestas del personal, entrevistas y las revisiones del servicio de extensión realizadas por el propio ministerio. La calidad y cantidad de los servicios entregados se evalúan usando una combinación de los hallazgos que arrojaron las evaluaciones participativas (de los beneficiarios), encuestas del personal, además de las medidas de la extensión y la naturaleza y frecuencia del contacto entre los agentes de extensión y los agricultores basados en la encuesta del Departamento de Evaluación de las Operaciones de 1998. Los datos de la encuesta también se usan para medir los resultados del programa, en términos de la conciencia de los agricultores y la aceptación de las recomendaciones hechas en el marco de la extensión.

Los resultados del programa, es decir, su real impacto sobre la producción agrícola en Kenya, se evalúan relacionando la provisión de servicios de extensión con los cambios en la productividad y la eficiencia a nivel de cada explotación agrícola. En base a los datos de panel de los hogares, estos impactos se estiman usando el análisis de datos del marco general, una técnica no paramétrica para medir los cambios en la eficiencia y productividad agrícola en el tiempo, junto con un análisis econométrico que mida el impacto de la prestación de servicios de extensión en la producción agrícola. Se usan métodos de valuación contingentes para averiguar directamente la voluntad de los agricultores de pagar por servicios de extensión.

IV. Resultados

El desarrollo institucional de NEP-I y NEP-II ha sido limitado. Después de 15 años, la eficacia de los servicios de extensión ha mejorado muy poco. Si bien últimamente se ha producido un correcto replanteamiento de las formas de abordar el tema de la extensión, en general el programa de extensión ha carecido de la visión estratégica para un futuro desarrollo. La administración del programa sigue siendo deficiente y prácticamente no existen sistemas de información. La calidad y cantidad de la prestación de servicios es pobre. Tanto los beneficiarios como el personal del servicio informan que las visitas son poco frecuentes y son poco eficaces. Si bien sigue habiendo una demanda no satisfecha por servicios técnicamente útiles, la atención del servicio público de extensión sigue puesta en mensajes agronómicos simples y básicos. Sin embargo, el planteamiento asumido (un contacto muy intenso con una cantidad restringida de agricultores) es adecuado para entregar información técnica. El resultado ha sido un sistema de prestación de servicios caro e ineficiente. Las actividades de extensión han tenido poca influencia sobre la evolución de patrones de conciencia y aceptación de las recomendaciones, lo que apunta a un limitado potencial de producir algún impacto. En términos del impacto real

sobre la producción y eficiencia agrícola, los datos apunan a un pequeño impacto positivo de los servicios de extensión sobre la eficiencia técnica, pero ninguno en la eficiencia económica general o la distribución. Más aún, usando los datos disponibles no es posible establecer un impacto significativo de la prestación de los servicios de extensión sobre la productividad a nivel de explotación agrícola. Sin embargo, los datos sí muestran que el impacto ha sido relativamente mayor en áreas anteriormente menos productivas, donde es probable que la falta de conocimientos haya sido también mayor. Estas conclusiones son coherentes con los hallazgos de la valuación contingente. Una amplia mayoría de agricultores, tanto receptores como no receptores del programa, están dispuestos a pagar por recibir asesoría, lo que indica que hay una demanda no satisfecha. Sin embargo, el valor percibido del servicio, en términos del monto ofrecido, está muy por debajo de lo que el gobierno está gastando actualmente en su prestación.

V. Implicancias para políticas

La Evaluación del Servicio de Extensión de Kenya se destaca por la cantidad de conclusiones de políticas prácticas que se pueden desprender de sus resultados, muchos de los cuales son pertinentes para el diseño de futuros proyectos de extensión agrícola. En primer lugar, en la evaluación se señala que es necesario mejorar la focalización de los servicios de extensión, concentrándose en áreas y grupos donde hay una mayor diferencia entre las prácticas promedios y adecuadas y, por lo tanto, el impacto debería ser mayor. Más aún, es necesario adaptar cuidadosamente la asesoría para satisfacer las demandas de los agricultores, tomando en cuenta las variadas condiciones tecnológicas y económicas locales. El lograr tener un nivel de servicio con estas características requiere un flujo regular y oportuno de información adecuada y confiable, al igual que un sistema de supervisión y evaluación que entregue retroalimentación regular sobre el contenido de los servicios desde los mismos beneficiarios-

Para aumentar la eficiencia del programa, los agentes de extensión deberían cubrir zonas más amplias, con una presencia menos intensa y frecuente, lo que probablemente sería más eficaz en función de los costos. No hay suficientes innovaciones técnicas que justifiquen una frecuencia muy alta de visitas y aquellos que aún no tiene acceso demandan servicios de extensión. El enfoque del programa para la entrega de servicios, que es de carácter muy general y depende fundamentalmente de una sola metodología (visitas a las explotaciones agrícolas) para entregar mensajes estándares simples, también restringe la eficiencia del programa. Los programas de radio gozan actualmente de popularidad, los agricultores más jóvenes son más educados y los proveedores alternativos (organizaciones no gubernamentales) están comenzando a surgir en Kenya rural. Una aproximación pluralista flexible para la entrega de servicios, particularmente una que aproveche medios de comunicación de menor costo, podría mejorar la eficacia del programa en función de los costos.

Por último, los principales resultados de la evaluación apuntan a la necesidad de una reforma institucional. Al igual que con otros servicios, se puede lograr una mayor eficiencia en la entrega de servicios de extensión con arreglos institucionales más adecuados. El enfoque central de la institución debería ser el cliente (agricultor). La

descentralización del diseño del programa, inclusive los mecanismos de participación que le entregan voz al agricultor (como compartir costos y organizaciones de agricultores), deben transformarse en una parte integral del mecanismo de entrega. La sustentabilidad es crítica. El tamaño e intensidad del servicio se debe basar en las brechas tecnológicas y de conocimiento actuales y en el ritmo de aparición de nuevas tecnologías. La recuperación de costos, aunque sea parcial, tiene varias ventajas: proporciona los incentivos adecuados, aborda temas de responsabilidad pública y control de calidad, hace que el servicio esté más orientado a la demanda y sea más sensible y proporciona cierto alivio presupuestario. Este tipo de arreglos institucionales permanecen aún sin conocerse en Kenya y en muchos programas de extensión de África y de todo el mundo.

VI. Costos y administración de la evaluación

Costos. El presupuesto total asignado a la evaluación fue de US\$250.000, lo que cubrió la recopilación y procesamiento de los datos de la encuesta domiciliaria (US\$65.000, si bien probablemente sea una subestimación de los gastos reales), encuesta del personal de extensión, datos e informe del consultor (US\$12.500), costos de recopilaciones de datos adicionales (US\$12.500) y un analista de investigación (US\$8.000). Se deben agregar aproximadamente US\$100.000 (que no se reflejan en los costos oficiales) por gastos de personal para el procesamiento y análisis de datos y la redacción de informes, para reflejar totalmente el costo del estudio.

Administración. Para mantener la objetividad y separar el trabajo de la encuesta tanto del servicio de extensión del gobierno como del Banco Mundial, la encuesta domiciliaria se llevó a cabo por el Tegemeo Institute de la Universidad de Egerton, un instituto de investigación independiente de Kenya. El análisis lo ejecutó Madhur Gautam del Banco Mundial.

VII. Lecciones aprendidas

- La combinación de una evaluación teórica y un marco basado en los resultados puede proporcionar una base sólida para evaluar el impacto de las intervenciones de proyectos, especialmente cuando hay muchos factores que podrían influir en los resultados esperados. El diseño de esta evaluación permitió medir indicadores clave en etapas críticas del ciclo del proyecto, vinculando la información del proyecto a los resultados esperados para recopilar evidencia suficiente del impacto.
- Una evaluación empírica exige una supervisión constante y profunda. Es posible simplificar considerablemente una evaluación si se cuenta con un sistema de supervisión y ponderación de buena calidad y funcionamiento, especialmente si se cuenta con datos básicos adecuados. Son escasas las ocasiones en que se cuenta con recursos apropiados para estas actividades. Esta evaluación también se benefició tremendamente del acceso a algunos datos (aunque limitados) de la etapa previa al proyecto y también de fuentes independientes de información para fines de comparación.
- La validación transversal de las conclusiones usando diferentes enfoques analíticos y fuentes de información es importante para recopilar un conjunto verosímil de

evidencias. La imperfección de los datos y problemas de implementación limitan el nivel de confianza con el cual cada método puede responder a preguntas de evaluación clave. Las ponderaciones cualitativas y cuantitativas se complementan fuertemente. La experiencia de esta evaluación específica indica que aún a falta de evaluaciones participativas de los beneficiarios, es posible incluir preguntas bien diseñadas en una encuesta para recopilar información cualitativa y cuantitativa. Dicha información puede proporcionar conocimientos útiles para complementar las evaluaciones cuantitativas.

- Si se aplica en forma adecuada, una evaluación contingente puede ser una herramienta útil, especialmente para ponderar el valor de un servicio público en funcionamiento. Los resultados de la aplicación de esta evaluación son estimulantes y las respuestas parecen ser racionales y razonables.

VIII. Fuentes

Banco Mundial, 1999. *World Bank Agricultural Extension Projects in Kenya: An Impact Evaluation*. Departamento de Evaluación de Operaciones, Informe N° 19523. Washington, D.C.

Además, los siguientes documentos de trabajo también están disponibles en el Departamento de Evaluación de Operaciones del Banco Mundial:

The Efficiency of the T&V system Agricultural Extension in Kenya: Results from a Household Survey

Awareness and Adoption of Extension Messages

Reconsidering the Evidence on Returns to T&V Extension in Kenya

Farmer Efficiency and Productivity Change in Kenya: An Application of the Data Envelopment Analysis

The Willingness to Pay for Extension Services in Kenya: An Application of the Contingent Valuation Method

Anexo 1.9: El impacto del Programa de Readiestramiento de México (PROBECAT) en el empleo y los salarios

I. Introducción

Este caso tiene características poco habituales, ya que el programa ha sido evaluado tres veces: primero por el Banco Mundial sobre la base de datos de 1992 (Revenge, Riboud y Tan, 1994), segundo por el Ministerio de Trabajo mexicano con datos de 1994 (STPS, 1995) y tercero, una actualización del Banco Mundial (Wodon y Minowa, 1999). Las metodologías que se emplearon en las dos primeras evaluaciones fueron bastante similares y arrojaron resultados también similares. Perfeccionamientos metodológicos aplicados a la tercera evaluación generaron resultados y conclusiones de política bastante diferentes. El hecho de que los resultados difieran sustancialmente entre las dos primeras y la tercera evaluación es un claro indicio de la importancia que reviste la metodología y los datos que se utilizan y de la prudencia con que se deben interpretar los resultados al evaluar un programa.

Descripción del proyecto. PROBECAT (Programa de Becas de Capacitación para Trabajadores) es un programa de capacitación mexicano a corto plazo que busca aumentar los ingresos y el empleo de los trabajadores desempleados y desplazados. PROBECAT se administra a través de las oficinas de empleo estatales. Los participantes reciben un salario mínimo durante el período de capacitación, el que dura entre uno a seis meses, y la oficina de empleo local proporciona la ubicación. El programa era pequeño en sus inicios (aproximadamente 50.000 participantes), pero ha crecido considerablemente en los últimos años hasta abarcar a más de 500.000 personas al año.

Puntos destacados de la evaluación. Se destaca lo siguiente:

- La evaluación de 1994 es interesante debido a cuatro motivos: la utilización imaginativa de los datos existentes, la construcción de un grupo de comparación pareado, el reconocimiento explícito de la naturaleza multifacética de los resultados de la intervención, especialmente para grupos heterogéneos de trabajadores, y el análisis costos-beneficios explícito. Las conclusiones de la evaluación fueron bastante positivas en términos del impacto del programa en los beneficiarios.
- La evaluación de 1995 es una réplica de la metodología utilizada en la evaluación de 1994, pero con un conjunto de datos más recientes. También aquí las conclusiones sobre el impacto del programa son favorables. Dado que el diseño y los resultados de la evaluación de 1995 coinciden con aquellos de la evaluación de 1994, esta evaluación no se analizará en este anexo.
- La evaluación de 1999 se ejecutó como parte del estudio de pobreza de México. Se usó el conjunto de datos utilizado para la evaluación de 1995, pero se aplicó una metodología econométrica diferente. Con los controles empleados para la endogeneidad de la participación en el programa se disipó el impacto del programa en la probabilidad de trabajar y en los salarios después de la capacitación. Si bien esto no significa que el programa carezca de beneficios, parece indicar que funciona más como una red de seguridad temporal para los desempleados que como un programa de capacitación.

II. Preguntas de la investigación y diseño de la evaluación

En la evaluación de 1994, los autores estimaron el impacto de la capacitación en (a) la probabilidad de tener empleo después de 3, 6 y 12 meses, (b) el tiempo para salir de la condición de desempleo, (c) el impacto en los ingresos mensuales, horas trabajadas a la semana y salario por hora y (d) la rentabilidad de la inversión.

En la evaluación de 1999 se examinaron las mismas preguntas, salvo las horas trabajadas a la semana y el salario por hora. Dado que en esa evaluación no se encontraron repercusiones en el empleo y los ingresos mensuales, la rentabilidad es cero, pero nuevamente es posible que el programa funcione como una red de seguridad.

El diseño de ambas evaluaciones es innovadora en la construcción del grupo comparativo. En ambos casos, las evaluaciones combinan una encuesta de panel existente sobre la fuerza laboral, la encuesta Nacional de Empleo (ENEU), con un panel de participantes en la capacitación para el mismo período. Aunque no existe una alternativa a esta combinación de encuestas debido a restricciones de datos, se puede criticar la construcción de la muestra conjunta (grupos de control y de tratamiento), como se analizó en la evaluación de 1999:

- Al utilizar los individuos desempleados de la ENEU para formar el grupo de control, se asume que ninguno de los individuos de la ENEU se benefició con el programa. Esto no es así, porque cada uno de los individuos de la ENEU tiene cierta probabilidad de haber participado en PROBECAT. Afortunadamente, en vista que el programa era de pequeña envergadura hasta 1993, es probable que sólo una minoría muy pequeña de individuos del grupo de control haya participado en el programa (los datos utilizados en la evaluación de 1999 corresponden a 1993-94);
- La combinación de dos muestras aleatorias (participantes en PROBECAT que están recibiendo capacitación e individuos sin empleo en la ENEU) no es una muestra aleatoria, de modo que al no contar con propiedades estándares para los residuos, los resultados de las regresiones podrían no arrojar estimaciones coherentes para los parámetros, especialmente en vista que los modelos utilizados no son sensibles a los supuestos de la normalidad bivariada. En ausencia de mejores datos, no es mucho lo que se puede hacer a este respecto.

Las principales diferencias entre las evaluaciones de 1994 y 1999 son las siguientes:

- En la evaluación de 1994, los autores intentaron abordar los problemas de sesgo de selección que generó la selección no aleatoria de participantes en PROBECAT estimando un modelo probit de la probabilidad de participar. Luego, el grupo de comparación se limita a aquellos individuos que tienen una elevada probabilidad de participar. En la evaluación de 1999, los autores sostienen que este método no elimina el problema de la endogeneidad. En su lugar, usan una variable instrumental variable para controlar por la endogeneidad de la participación en el programa.

- En la estimación de los ingresos de la evaluación de 1994, si bien se controla por la participación en PROBECAT, no se considera el sesgo de selección de la muestra que resulta de la decisión de trabajar. En el estudio de 1999 se consideran ambos problemas de selección de la muestra.

III. Datos

En la evaluación de 1994, la información sobre los individuos que recibieron capacitación se recopiló de una encuesta retrospectiva de 1992 que se aplicó a 881 hombres y 845 mujeres que recibieron capacitación en 1990. Esto se complementa con los datos de panel sobre 371 hombres y 189 mujeres derivados de una encuesta domiciliaria aplicada en las 16 principales áreas urbanas de México. Esta encuesta formó parte de una encuesta trimestral regular de la fuerza de trabajo, ENEU, realizada por la agencia estadística mexicana. Los autores aprovecharon la estructura de rotación de grupos de la encuesta para tomar a los trabajadores que estuvieron sin empleo durante el tercer trimestre de 1990 y luego hacerles un seguimiento durante un año. Esto se complementó con un grupo que quedó sin empleo durante el cuarto trimestre de ese mismo año, haciéndoles un seguimiento durante nueve meses. En la evaluación de 1999 se utilizó este mismo método, pero con datos más recientes.

IV. Técnicas econométricas

Las técnicas econométricas utilizadas son análisis de supervivencia (modelos de duración) para la probabilidad de trabajar y regresiones Heckman para los salarios. Los siguientes análisis se basan en la evaluación de 1999. Se destacan las diferencias con la evaluación de 1994.

Impacto del PROBECAT en la duración de la búsqueda de empleo. En el análisis de supervivencia, la función de supervivencia $S(t)$ representa la duración del desempleo después de la capacitación (medida en meses). Dado $S(t)$, la función de riesgo $\lambda(t)$ que denota la posibilidad de quedar sin empleo (o el riesgo de seguir desempleado) en el momento t entre individuos que todavía no están empleados en ese momento es $\lambda(t) = -d(\log S(t))/dt$. La curva de supervivencia se puede especificar como una función de la participación en el programa P , características individuales X y características del estado Z , de modo que $\lambda = \lambda(t; X, Z, P)$. En el modelo de riesgo proporcional de Cox, si i denota un hogar y j denota el área de residencia del hogar, tenemos:

$$\lambda(t; X, Z, P1, P2) = \lambda_0(t) \exp(\gamma' X_{ij} + \delta' Z_j + \mu P_{ij}) \quad (1)$$

Cox propuso una estimación de probabilidad máxima parcial de este modelo, en que no se necesita especificar la función básica $\lambda_0(t)$. Si μ es positivo y estadísticamente significativo, el programa afecta positivamente el empleo. De un modo estilizado, la diferencia entre las evaluaciones de 1994 y 1996 se puede describir de la siguiente manera:

- En la evaluación de 1994, los autores ejecutaron una probit (unidad de probabilidad) de la participación en el programa y eliminaron a aquellos individuos del grupo de control que tenían una baja probabilidad de participar en el programa. A continuación ejecutaron la ecuación (1) sin controlar adicionalmente por la endogeneidad.
- En la evaluación de 1999, los autores también ejecutaron una probit de la participación en el programa, pero usaron la disponibilidad del programa en el nivel local (que obtuvieron de datos administrativos) como determinante adicional de la participación (pero no del resultado supeditado a la participación individual). Luego ejecutaron la ecuación (1) no con el valor efectivo de la variable de participación, sino con el valor (de índice) predicho obtenido de la probit de primera etapa. Se trata de un procedimiento de variable instrumental. La idea sigue los lineamientos de un trabajo sobre la evaluación de programas usando propiedades de descentralización realizado por Ravallion y Wodon (2000) y Cord y Wodon (1999). Los autores comparan sus resultados con otros métodos, demostrando que otros métodos muestran un sesgo en el valor de las estimaciones de parámetros debido a un insuficiente control de la endogeneidad.

Impacto de PROBECAT en los ingresos mensuales. Para realizar este análisis, en la evaluación de 1999 se utiliza un modelo con controles de la selección de muestras de la participación en la fuerza laboral y el programa (en la evaluación de 1994 sólo se controla por la participación en el programa). El logaritmo del salario esperado de un individuo se denota mediante el logaritmo w . Este salario es no cero exclusivamente si es mayor que el salario de reserva del individuo (de lo contrario, el individuo optaría por no trabajar). La diferencia no observada en el salario no esperado del individuo y su salario de reserva se denota por Δ^* . El salario esperado del individuo está determinado por un número de variables del individuo (vector E , que consta esencialmente de la educación y experiencia pasada del individuo) y geográficas Z , más la participación en el programa P . La diferencia entre el salario esperado del individuo y su salario de reserva es determinada por las mismas variables, más el número de hijos, el hecho de ser jefe/a de hogar y el hecho de estar casado, capturado por D . Por consiguiente, el modelo es:

$$\Delta_{ij}^* = \phi_{\Delta}'E_{ij} + \pi_{\Delta}'D_{ij} + \eta_{\Delta}'Z_j + \alpha_{\Delta}P_{ij} + v_{ij} \text{ con } \Delta_{ij} = 1 \text{ si } \Delta_{ij}^* > 0, \text{ y } 0 \text{ si } \Delta_{ij}^* < 0 \quad (2)$$

$$\text{Log } w_{ij}^* = \phi_w'E_{ij} + \eta_w'Z_j + \alpha_w P + \kappa_{ij} \text{ con } \text{Log } w = \text{log } w^* \text{ si } \Delta=1 \text{ y } 0 \text{ si } \Delta=0 \quad (3)$$

En la evaluación de 1999, del mismo modo que en el modelo de supervivencia, para controlar por la endogeneidad de la participación en el programa primero se estima la probit para la participación en el programa usando la disponibilidad en el nivel local como determinante de la participación individual. A continuación se estiman las ecuaciones recién indicadas usando el valor predicho (de índice) de la participación en el programa en lugar de su valor verdadero. En la evaluación de 1994, el modelo no controla por la decisión de participar en el mercado laboral que se entrega en la ecuación (2) más arriba. Esta ecuación es reemplazado por la probit de la participación en el programa, que se estima sin la disponibilidad local del programa a modo de variable

independiente. Nuevamente, las comparaciones de diversos modelos indican que hay un sesgo cuando no se utiliza la técnica de la variable instrumental.

V. Ejecutores

La evaluación de 1994 fue llevada a cabo por Ana Revenga, del Departamento de Países II de América Latina y el Caribe del Banco Mundial, Michelle Riboud del Departamento de Países IV de Europa y Asia Central del Banco Mundial y Hong Tan del Departamento de Desarrollo del Sector Privado del Banco Mundial. La evaluación de 1999 fue realizada por Quentin Wodon y Mari Minowa, también del Banco Mundial (región de América Latina.)

VI. Resultados

Los resultados que se obtuvieron en las diversas evaluaciones difieren mucho. En las evaluaciones de 1994 y 1995 se determinó que el programa tuvo impactos positivos en el empleo y los salarios. No se encontró impacto positivo alguno en la evaluación de 1999, la que se basa en los mismos datos utilizados para la evaluación de 1995. En términos del análisis de costos-beneficios, las primeras dos evaluaciones son favorables, pero la última no lo es. Los decepcionantes resultados de esta última evaluación no deben sorprender. Según estudios, los impactos de la mayoría de los programas de readiestramiento realizados en países de la OCDE han sido limitados y en el caso de los programas que han tenido algún impacto, éste suele desaparecer después de unos pocos años. (Dar y Gill, 1998). Sin embargo, el hecho de que PROBECAT podría no ser beneficioso para los participantes en el mediano a largo plazo de acuerdo a la última evaluación no significa que deba ser eliminado. Más que una capacitación, el programa se podría considerar como una red de seguridad temporal (a través de la remuneración del salario mínimo). O se podría mejorar de modo que preste una capacitación de impactos más duraderos.

VII. Lecciones aprendidas

Además de algunas de las características innovadoras de estas evaluaciones y sus limitaciones, la lección clave es que las evaluaciones de los programas y la utilización de los resultados para recomendar opciones de política se deben abordar con extremo cuidado. Siempre se debe tener presente que una evaluación posterior que utilice técnicas econométricas diferentes puede contradecir los resultados de una anterior. En la literatura se conocen muchos casos de este tipo.

VIII. Fuentes

Revenga, Ana, Michelle Riboud y Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages". *World Bank Economic Review* 8(2): 247-77.

Wodon, Quentin, y Mari Minowa, "Training for the Urban Unemployed: A Reevaluation of Mexico's PROBECAT". Banco Mundial, Government Programs and Poverty in Mexico, Informe N° 19214-ME, Volumen II.

Anexo 1.10: México, Programa Nacional de Educación, Salud y Nutrición: (PROGRESA)

Propuesta para una evaluación

I. Introducción

Descripción del proyecto. PROGRESA es un programa multisectorial destinado a combatir la extrema pobreza en México mediante la entrega de un paquete integral de servicios de salud, nutrición y educación a familias pobres. Durante al menos tres años consecutivos, el Gobierno de México proporcionará asistencia monetaria, suplementos nutricionales, subvenciones educacionales y un paquete de salud básica. Además, tiene la intención de ampliar PROGRESA de su tamaño actual de 400.000 familias a 1-1,5 millones de familias hacia fines de 1998, con un gasto de \$500 millones.

Puntos destacados de la evaluación. La evaluación es especialmente compleja, ya que se evalúan tres dimensiones del programa: operación, eficacia de la orientación hacia beneficiarios específicos e impacto. La multidimensionalidad de los resultados mismos aumenta aun más su complejidad. Por consiguiente, la evaluación abarca muchos componentes diferentes: selección de beneficiarios, métodos de evaluación, marco analítico no experimental, necesidades de datos, impactos en la educación, en la salud, en el consumo de alimentos y la nutrición, en los gastos de consumo y su asignación intrafamiliar, posibles impactos de una segunda ronda del programa, simulaciones de cambios en los beneficios del programa y aspectos de eficacia en función de los costos y de costos-beneficios.

Si bien la evaluación constituye más bien un esbozo de ideas que los resultados de una puesta en práctica, una importante lección que se aprende de ella es cómo considerar cuidadosamente una evaluación y estructurarla antes de implementarla efectivamente. Específicamente, hay un bosquejo muy útil de los aspectos conceptuales y empíricos que se deben abordar en una evaluación y las formas en que es posible enfrentarlos. Otro componente útil de la evaluación es su amplitud: más que evaluar simplemente el impacto de una intervención, ayudará a indicar exactamente si los resultados se deben a éxitos o fracasos en la operación y orientación hacia beneficiarios específicos de la intervención.

II. Preguntas de la investigación y diseño de la evaluación

Las preguntas centrales de la investigación están destinadas a evaluar las tres dimensiones del desempeño de PROGRESA, esto es, aspectos operativos, orientación hacia beneficiarios específicos e impacto. Con frecuencia se ignora el aspecto operativo de una intervención, a pesar del hecho de que al adoptar las medidas correctivas se podría convertir el fracaso de una intervención en éxito. Lo mismo se puede sostener respecto a la orientación hacia beneficiarios específicos: puede parecer que un programa ha fallado simplemente debido a su deficiente focalización, más que debido a un defecto mismo de

la intervención. La evaluación del impacto es más estándar, aunque incluso esta meta es bastante ambiciosa, ya que se analiza tanto la magnitud del impacto como los medios a través de los cuales se alcanza.

El funcionamiento del programa se supervisa en un procedimiento de dos etapas. El equipo desarrolla un esquema secuencial de los pasos para la intervenciones. Luego se basa en observaciones, entrevistas, grupos representativos (*focus groups*) y talleres con partes interesadas para evaluar, analizar y eventualmente cambiar los procesos del programa.

También se utiliza un método de dos etapas para orientar PROGRESA hacia hogares beneficiarios específicos. El primer paso es identificar, mediante un índice basado en la pobreza, las localidades dentro de una región que cumplen con los requisitos para recibir el programa. El segundo paso es identificar la elegibilidad de una familia dentro de la localidad, sobre la base de la interacción entre funcionarios de PROGRESA y dirigentes locales. Para abordar la validez de esta focalización, en el estudio (a) se compara la distribución de los niveles de consumo entre los hogares participantes y no participantes dentro de las localidades donde se aplica el programa, (b) se deriva un punto de corte para la elegibilidad del consumo de los hogares que sea coherente con el número total de hogares a los que PROGRESA puede atender, (c) se realiza un análisis de sensibilidad y de especificidad de los hogares con PROGRESA y sin PROGRESA frente los hogares seleccionados y no seleccionados bajo este punto de corte, (d) se explora la capacidad de los actuales criterios para predecir el consumo, (e) se identifican criterios alternativos de otras fuentes de datos y (f) se simulan modelos que podrían mejorar la orientación hacia beneficiarios específicos con criterios alternativos (Instituto Internacional de Investigaciones sobre Políticas Alimentarias, 1998, pág. 6)

Para evaluar el impacto se usó el mismo sistema, lo que generó una asignación aleatoria de las localidades a 296 grupos de tratamiento y 173 grupos sin tratamiento, con 14.382 familias en la primera categoría y 9.202 familias en la segunda. Las familias en la categoría de control que cumplen con los requisitos recibirán el tratamiento del programa una vez transcurrido por lo menos un año.

Los consultores piensan examinar una posible no aleatorización comparando las características de los grupos de tratamiento y de control. Si son sistemáticamente diferentes, se usarán tres métodos no experimentales: métodos de la función de control, métodos de correlación y métodos de regresión.

III. Datos

El componente de los datos operativos se obtiene de observaciones y entrevistas, grupos representativos y talleres con partes interesadas. Se centra principalmente en identificar las cosas que están pasando y sus motivos, el nivel de satisfacción con el proceso y sugerencias para mejorarlo. Esta información se recopila en todas las localidades y también se basará en gran medida en los registros administrativos internos de PROGRESA.

Se han implementado dos encuestas: las encuestas del censo de 1997 y la encuestas

básicas de marzo de 1998. Claramente, la variable central para el criterio de focalización es el consumo de los hogares. Si bien esta información no fue recopilada en el censo, sí lo fue en la encuesta de marzo. Sin embargo, esta variable no incluye información sobre el autoconsumo y si bien esta información se recopilará más adelante, estará contaminada por la aplicación de PROGRESA. Los consultores tienen la intención de trabajar exclusivamente con los hogares elegibles y no elegibles en las localidades de control.

La evaluación del impacto depende de la selección de los respectivos indicadores. PROGRESA debería tener un impacto sobre la calidad y la cantidad de los servicios prestados, como también sobre la inversión en salud, nutrición y educación. Se proponen un sinnúmero de indicadores de evaluación basados en varios resultados del impacto y cada uno tiene una fuente de datos asociada. Por medio de encuestas básicas y de seguimiento se medirá el bienestar del hogar (en términos del consumo familiar), ahorros, acumulación de bienes duraderos; con una encuesta básica de submuestras de nutrición y encuestas de seguimiento se medirá el estado nutricional de salud de los niños; el progreso educacional de los niños se medirá mediante pruebas nacionales normalizadas; el consumo de alimentos se captará por medio de las encuestas básica y de seguimiento; la utilización de las escuelas se abordará tanto mediante una encuesta a nivel de escuela como por las encuestas básica y de seguimiento; el uso de los establecimientos de salud se puede supervisar a través de expedientes clínicos y encuestas; y la condición de las mujeres también se puede medir por medio de encuestas y de investigaciones de las partes interesadas.

Una característica muy atractiva de la evaluación que se propone es el enfoque analítico adoptado para examinar las medidas de los resultados actuales y el análisis en gran escala de medidas de resultados y control más apropiadas para la educación, la salud y el consumo.

Se pretende realizar un análisis de costos-beneficios. Se desarrolla un conjunto de beneficios, no obstante la dificultad inherente a la monetización de los mejoramientos en la calidad de vida y el empoderamiento. También se identifican dos diferentes tipos de costos: costos administrativos del programa y costos del programa. Los primeros incluyen los costos de selección, mecanismos de prestación focalizada y supervisión y los segundos incluyen la generación de ingresos sacrificados.

IV. Técnicas econométricas

Las técnicas econométricas aplicadas dependen de las relaciones que se deben estimar. Los consultores analizan la conveniencia de la relación producción-función (por ejemplo, para el progreso académico), relaciones de demanda (por ejemplo, para los servicios de salud y educación) y relaciones de demanda condicional (en que algunas variables son más bien determinadas por la familia y no por el individuo).

La técnica econométrica más interesante que se usó se aplica a la estimación de una función de gasto de trabajo Leser, de la forma

$$W_j = \alpha_1 + \beta_{1j} \text{lpccexp} + \beta_{2j} \text{lsiz} + \sum_k \delta_{kj} \text{dem}_k + \sum_s \Theta_{sj} z_s + \beta_{3j} P + e_j,$$

donde w_j es la proporción del bien j th en el presupuesto; $\ln pc_{exp}$ es el logaritmo del gasto per cápita; $\ln sz$ es el logaritmo del tamaño del hogar; dem_k es la proporción del grupo demográfico k en el hogar; z_s es un vector de variables ficticias que afectan la ubicación del hogar; P capta la participación en ProgresA y e_j es el término de error.

Este planteamiento tiene muchas ventajas: permite la inclusión de factores de control; satisface la restricción de la adición; y es de uso generalizado, lo que permite hacer comparaciones con otros estudios. Finalmente, el modelo se puede usar para identificar tres trayectorias diferentes en que PROGRESA puede influir en los gastos: cambiando los recursos de los hogares (β_{1j} veces la propensión marginal a consumir, estimada de manera separada); cambiando la distribución del ingreso (mediante su modificación para que incluya la proporción de mujeres adultas en el hogar) y a través de un mayor impacto de la participación. Las encuestas básica y de seguimiento permiten usar metodologías de diferencia en las diferencias.

También se identifican los aspectos econométricos clave que probablemente se deberán enfrentar: colinearidad, error de medición, variables omitidas, simultaneidad, además de la identificación de un período de tiempo dentro del cual es razonable esperar que se pueda observar un impacto.

V. Ejecutores

El personal del Instituto Internacional de Investigaciones sobre Políticas Alimentarias incluye a Gaurav Datt, Lawrence Haddad, John Hoddinott, Agnes Quisumbing y Marie Ruel. El equipo incluye a Jere Behrman, Paul Gertler y Paul Schultz.

VI. Lecciones aprendidas

La principal lección que se aprende aquí es el valor de identificar los problemas, metodología y fuentes de datos de una evaluación y de evaluar críticamente esta evaluación antes de realizarla. Este esbozo de la evaluación presta un servicio muy valioso al preparar una ilustración cuidadosa de todos los posibles problemas y escollos que podría encontrar un evaluador. Específicamente, se identifican algunos asuntos de sentido común que se presentan al evaluar un impacto: (a) podría ser difícil predecir los cambios de políticas debido a sustituciones cruzadas y ajustes de comportamiento, (b) los beneficios y costos marginales dependen de diversos aspectos: externalidades (poniendo una brecha entre la valoración social y privada), los actores (padres versus niños), (c) la importancia de las características no observadas, (d) la importancia de controlar por las características individuales, familiares y de la comunidad y (e) la dependencia de las estimaciones empíricas de un determinado entorno macroeconómico, normativo, de mercado y de políticas.

VII. Fuente

Instituto Internacional de Investigaciones sobre Políticas Alimentarias. 1998. Programa Nacional de Educación, Salud y Alimentación (PROGRESA): *A Proposal for Evaluation* (con apéndice técnico). Washington, D.C.: IFPRI

Anexo 1.11: Evaluación de la reforma escolar de Nicaragua: Un planteamiento cuantitativo y cualitativo combinado

I. Introducción

Descripción del proyecto. En 1991, el Gobierno de Nicaragua implementó una reforma de amplio alcance para su sistema de educación pública. En el marco de este proceso de reforma se descentralizó la administración de las escuelas (decisiones sobre personal, presupuestos, planes de estudio y aspectos pedagógicos) y se traspasaron las responsabilidades financieras al nivel local.

Las reformas se incorporaron gradualmente a partir de un decreto de 1991 que estableció los consejos comunitarios de padres en todas las escuelas públicas. Luego, en el marco de un programa piloto aplicado en 1993 en 20 escuelas secundarias seleccionadas cuidadosamente, estos consejos se transformaron en consejos administrativos de escuelas con mayores atribuciones en materia de personal, presupuestos, planes de estudio y aspectos pedagógicos. En 1995, estos consejos estaban funcionando en 100 escuelas secundarias y más de 300 escuelas primarias, las que se incorporaron al programa a través de un proceso de autoselección que implicaba una solicitud de los maestros y directores de escuela. Se espera que la autonomía de las escuelas sea prácticamente universal hacia fines de 1999.

La meta de las reformas nicaragüenses es mejorar el aprendizaje de los estudiantes a través de la modificación de los procesos de organización dentro de las escuelas públicas, de modo que la principal prioridad al tomar decisiones sea el bienestar de los estudiantes. A medida que la administración de las escuelas sea más democrática y participativa y aumenten los ingresos generados en el nivel local, los patrones de gasto serán más racionales y asignados a esfuerzos que mejoren directamente los aspectos pedagógicos y eleven el aprovechamiento escolar.

Evaluación del impacto. La evaluación de la Reforma nicaragüense de autonomía en las escuelas representa uno de los primeros esfuerzos sistemáticos por el evaluar el impacto de la descentralización de las escuelas en los resultados de los estudiantes. La evaluación, realizada conjuntamente por el Banco Mundial y el Ministerio de Educación, se inició en 1995 y debía concluir hacia fines de 1999. Lo novedoso de su diseño es que combina tanto métodos de evaluación cualitativos como cuantitativos. Es excepcional en el aspecto del componente cuantitativo, en cuanto incluye un módulo separado para evaluar los procesos de toma de decisiones en las escuelas. La evaluación también es un ejemplo de técnicas de 'prácticas óptimas' cuando no se cuenta con datos de referencia (básicos) y cuando una aplicación selectiva (no aleatoria) de las reformas excluye la posibilidad de un diseño de evaluación experimental.

Con el componente cualitativo de la evaluación se busca determinar si en las escuelas efectivamente se están observando las presuntas reformas administrativas y financieras y

evaluar la forma en que las diferentes partes interesadas consideraban el proceso de reforma. El componente cuantitativo sirve para dimensionar estos resultados al responder a la siguiente pregunta: “Los cambios en la administración y financiamiento de las escuelas ¿realmente mejoran los resultados de aprendizaje de los niños?” Los resultados cualitativos indican que una implementación exitosa de las reformas depende en gran medida del contexto y entorno de las escuelas (por ejemplo, nivel de pobreza de la comunidad), mientras que los resultados cuantitativos apuntan a que una mayor toma de decisiones por parte de las escuelas está efectivamente asociada de manera significativa con un mejor rendimiento de los estudiantes.

II. Diseño de la evaluación

El diseño de la evaluación de la Reforma nicaragüense de autonomía en las escuelas se basa en la ‘técnica de comparación pareada’. De acuerdo con ésta, los datos de una muestra representativa de escuelas que participan en el proceso de reforma se comparan con datos de una muestra de escuelas que no participan en él. La muestra de las escuelas no participantes se selecciona de modo que éstas coincidan lo más exactamente posible con las características de las escuelas participantes, proporcionando así el escenario contrafactual. Se escogió este diseño, porque la falta de datos básicos o de referencia impidió el uso de una técnica de evaluación de ‘antes’ y ‘después’, y porque las reformas no se aplicaron aleatoriamente a las escuelas, lo que excluyó la posibilidad de usar un diseño de evaluación experimental (en que la muestra de las escuelas examinadas en la evaluación sería aleatoria y por consiguiente representativa a nivel nacional).

III. Técnicas de recopilación y análisis de datos

El estudio cualitativo se basa en los datos de una muestra de 12 escuelas, 9 de ellas sometidas al proceso de reforma y las 3 restantes no, las que representan el grupo de control (de hecho, se recopilaron datos de 18 escuelas, pero de éstas sólo se incluyeron 12 en el estudio cualitativo debido a la demora en lograr la preparación de las transcripciones y a una decisión de concentrar el grueso del análisis en las escuelas reformadas, lo que proporcionó material más pertinente para el análisis). La muestra de 12 escuelas se escogió de modo que representara a escuelas primarias y secundarias, rurales y urbanas y, con la ayuda de datos de la encuesta cuantitativa de 1995, tuvieran diferentes grados de efectiva autonomía en la toma de decisiones. Se realizaron un total de 82 entrevistas y sesiones con grupos representativos, centrándose en descubrir la manera en que los directores de escuelas, miembros de los consejos, padres y apoderados entendían y visualizaban el proceso de descentralización. Todas las entrevistas fueron realizadas por nicaragüenses oriundos del país, capacitados a través de simulaciones de entrevistas y pruebas piloto para usar una serie de preguntas dirigidas sin inducir las respuestas. Las entrevistas se grabaron en cinta, se transcribieron y luego se condensaron en una transcripción de dos a cuatro páginas. A continuación, estas páginas se analizaron para identificar conjuntos discretos de evidencia y temas fundamentales que fueran comunes a las escuelas y actores y a las escuelas reformadas y el grupo de control.

La recopilación de datos cuantitativos tuvo dos componentes, una encuesta de panel de escuelas que se realizó en dos tandas (noviembre - diciembre de 1995 y abril - agosto de 1997), y pruebas del rendimiento escolar para estudiantes de estas escuelas que se realizaron en noviembre de 1996. En la encuesta de las escuelas se recopiló información sobre tasas de matrícula, repetición y deserción, recursos físicos y humanos, toma de decisiones a nivel de escuela y características de los directores, maestros, estudiantes y sus familias. El módulo de toma de decisiones a nivel de escuela es único en su tipo y presenta una serie de 25 preguntas destinadas a evaluar sí y en qué medida la reforma ha aumentado efectivamente la toma de decisiones de parte de las escuelas. La encuesta abarcó a 116 escuelas secundarias (73 reformadoras y 43 no reformadoras que representan el grupo de control) y 126 escuelas primarias (80 reformadoras y 46 no reformadoras). También aquí los grupos de control se seleccionaron de modo que coincidieran con las características de las escuelas reformadas. En la encuesta también se recopilaron datos sobre 400 maestros, 182 miembros de consejo y 3.000 estudiantes y sus padres, escogiendo en forma aleatoria entre 10 a 15 estudiantes de cada escuela. Se aplicaron pruebas de rendimiento a fines del año escolar 1996 a los estudiantes que permanecieron en la escuela y que podían seguir, y nuevamente en 1997 durante la segunda ronda de recopilación de datos de la encuesta.

El análisis de los datos cuantitativos se basa en técnicas de regresión para estimar una función de producción de la educación. Con esta técnica se examinan las repercusiones del sistema administrativo de la escuela (su grado de descentralización) en los niveles de rendimiento de los estudiantes, controlando por los insumos escolares y las características de los hogares y estudiantes. El análisis mide el impacto de la descentralización '*de jure*' y '*de facto*'; la descentralización *de jure* sólo indica si la escuela se ha integrado o no válidamente a la reforma, mientras que la descentralización *de facto* mide el grado de la autonomía efectiva que ha alcanzado la escuela. La descentralización *de facto* se mide como el porcentaje de las 25 decisiones clave que realiza la escuela misma y se espera que varíe entre las escuelas debido a la introducción gradual de las reformas (de modo que las escuelas de la muestra se encontrarán en diferentes etapas del proceso de reforma) y porque hay distintas capacidades para implementar exitosamente las reformas de acuerdo al contexto específico de la escuela (un resultado que se identificó en el estudio cualitativo).

IV. Resultados

El estudio cualitativo indica que los cambios de políticas en el nivel central no siempre generan flujos causales regulares hacia el nivel local. Las reformas en general están asociadas con una mayor participación de los padres y con mejoramientos en la administración y las funciones rectoras. Pero el grado de éxito en la aplicación de las reformas fluctúa de acuerdo al contexto de la escuela. Especialmente importante es el grado de empobrecimiento de la comunidad que la rodea (resulta difícil aumentar el financiamiento local en las escuelas de comunidades pobres) y el grado de cohesión entre el personal de la escuela (donde el éxito de la descentralización ha sido limitado cuando actores clave como los maestros no se sienten integrados al proceso de reforma). Los gestores de políticas a menudo ignoran la gran variabilidad de los contextos locales en los que se introducen los nuevos programas. Los resultados cualitativos apuntan a que en la

práctica la meta de un mayor financiamiento local de las escuelas probablemente no es factible, especialmente en las comunidades pobres, y que por consiguiente, es un aspecto que se debe replantear.

El estudio cuantitativo fortalece la conclusión de que las escuelas reformadas de hecho están tomando más decisiones por su cuenta, especialmente en materia de asuntos pedagógicos y de personal. La autonomía *de jure*, esto es, si una escuela ha firmado o no el contrato de reforma, no necesariamente se traduce en una mayor toma de decisiones a nivel de la escuela, ni tampoco afecta del mismo modo a todas las escuelas. El grado de autonomía alcanzado depende del nivel de pobreza de la comunidad y del tiempo que la escuela ha estado participando en el proceso de reforma. Los resultados de las regresiones indican que la autonomía *de jure* influye poco en los resultados de rendimiento de los estudiantes; pero la autonomía *de facto*, el grado de descentralización efectiva alcanzado por la escuela, sí está significativamente relacionada con un mejor rendimiento de los estudiantes (se trata de un resultado preliminar mientras se exploran adicionalmente los datos de panel que han quedado disponibles recientemente). Asimismo, las simulaciones indican que una mayor descentralización de las escuelas influye más en el rendimiento de los estudiantes que los mejoramientos de otros indicadores en que típicamente se centran las políticas, como capacitación de los maestros, menor número de estudiantes por curso y un mayor número de libros de texto, etc.

V. Aplicación de políticas

Los resultados de la evaluación proporcionan evidencia concreta sobre los resultados tangibles que ha arrojado la Reforma nicaragüense de autonomía de en escuelas. Las escuelas reformadas efectivamente están tomando más decisiones a nivel local (la descentralización está ocurriendo realmente, no sólo en los libros) y la mayor toma de decisiones local sí resulta en un mejor rendimiento de los estudiantes.

Los resultados también destacan varias áreas en que se pueden mejorar las políticas. Debido a ello, el Ministerio de Educación ha introducido varios cambios al programa de reforma escolar. El programa ahora pone mayor énfasis en la función de los maestros y en promover los aspectos pedagógicos de la reforma. La capacitación de los maestros se ha incluido como parte del programa y se está considerando el establecimiento de un Consejo pedagógico. Asimismo, en respuesta a los problemas de financiamiento de las comunidades pobres, el Ministerio está elaborando un esquema de subsidios basado en el mapa de la pobreza. Finalmente, los beneficios tangibles de esta evaluación han impulsado al Ministerio a incorporar un componente de evaluación permanente en el programa de reforma.

VI. Costos y administración de la evaluación

Costos. El costo total de la evaluación sumó aproximadamente \$495.000, lo que representa menos del 1,5% del préstamo del Banco Mundial (este total no incluye el costo de los equipos locales de contrapartida del Ministerio de Educación nicaragüense). De este costo total de la evaluación, el 39% se gastó en soporte técnico proporcionado por

consultores externos, el 35% en la recopilación de datos, el 18% en tiempo del personal del Banco Mundial y el 8% restante en viajes.

Administración. La evaluación fue realizada conjuntamente por el Ministerio de Educación nicaragüense y el Banco Mundial. En Nicaragua, el equipo de evaluación estuvo encabezado por Patricia Callejas, Nora Gordon y Nora Mayorga de Caldera del Ministerio de Educación. En el Banco Mundial, la evaluación se ejecutó como parte del proyecto de investigación “*Impact Evaluation of Education Projects Involving Decentralization and Privatization*” (Evaluación de impacto de proyectos educacionales que implican descentralización y privatización) bajo la dirección de Elizabeth King, con Laura Rawlings y Berk Ozler. Bajo la coordinación del equipo del Banco Mundial, Bruce Fuller y Magdalena Rivarola de Harvard School of Education colaboraron con Liliam López del Ministerio de Educación nicaragüense para realizar la evaluación cualitativa.

VII. Lecciones aprendidas

Valor del enfoque de método combinado. Al usar técnicas de investigación cualitativas y cuantitativas, surgió una combinación de resultados útiles y pertinentes para las políticas. El trabajo cuantitativo proporcionó una visión general amplia y estadísticamente válida de las condiciones y resultados de las escuelas; el trabajo cualitativo perfeccionó estos resultados, permitiendo comprender mejor por qué algunos de los resultados esperados de la reforma fueron exitosos, mientras que otros fallaron, lo que sirvió de orientación para ajustar las políticas. Asimismo, debido a que es más intuitivo, el trabajo cualitativo fue más accesible y en consecuencia interesante para el personal del ministerio, lo que a su vez ayudó a otorgar credibilidad al proceso de evaluación y a construir rápidamente capacidades dentro del ministerio.

Importancia de la creación de capacidades locales. La creación de capacidades locales fue onerosa y exigió frecuentes contactos y coordinación con las contrapartidas del Banco Mundial y los consultores externos. Sin embargo, su beneficio fue el rápido desarrollo de un sentido de pertenencia y responsabilidad local frente al proceso de evaluación, lo que a su vez estimuló un alto grado de aceptación de los resultados de la evaluación (ya sea que éstos se reflejaran positiva o negativamente en el programa). Estos resultados de la evaluación sirvieron de aporte directo a la reforma a medida que se fueron produciendo. El impacto de la evaluación en las políticas también se vio incrementado por un equipo local cohesionado, en que los evaluadores y gestores de políticas trabajaron en colaboración, y porque el Ministro de Educación se integró al proceso de evaluación como un defensor integral.

VIII. Fuentes

En los siguientes documentos se entrega información detallada sobre la evaluación de la Reforma nicaragüense de autonomía en las escuelas:

Fuller, Bruce, y Magdalena Rivarola. 1998. *Nicaragua's Experiment to Decentralize Schools: Views of Parents, Teachers and Directors*. Serie de documentos de trabajo sobre

Evaluación del impacto de las reformas educacionales, documento N° 5. Banco Mundial, Washington, D.C.

King, Elizabeth, y Berk Ozler. 1998. *What's Decentralization Got To Do With Learning? The Case of Nicaragua's School Autonomy Reform*. Serie de documentos de trabajo sobre Evaluación del impacto de las reformas educacionales, documento N° 9. Banco Mundial, Washington, D.C.

King, Elizabeth, Berk Ozler y Laura Rawlings. 1999. *Nicaragua's School Autonomy Reform: Fact or Fiction?* Washington, D.C.: Banco Mundial.

Equipo de evaluación de la reforma nicaragüense. 1996. *Nicaragua's School Autonomy Reform: A First Look*. Serie de documentos de trabajo sobre Evaluación del impacto de las reformas educacionales, documento N° 1. Banco Mundial, Washington, D.C.

Equipo de evaluación de la reforma nicaragüense. 1996. *1995 and 1997 Questionnaires, Nicaragua School Autonomy Reform*. Serie de documentos de trabajo sobre Evaluación del impacto de las reformas educacionales, documento N° 7. Banco Mundial, Washington, D.C.

Rawlings, Laura. 2000. "Assessing Educational Management and Quality in Nicaragua." En Bamberger, *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C.: Banco Mundial.

Anexo 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

I. Resumen de la evaluación

Los recursos para la educación son extremadamente restringidos en la mayoría de los países pobres, de modo que es importante asignarlos de manera eficiente. De las tres opciones comunes de políticas disponibles: clases de menor tamaño, programas de capacitación más prolongados para los maestros y la entrega de libros de texto, sólo esta última opción con frecuencia ha mostrado un impacto positivo significativo en el aprendizaje de los estudiantes. En esta evaluación se cuantificó el impacto de la disponibilidad de libros de texto en el aprendizaje de matemáticas de los estudiantes nicaragüenses de primer grado.

El diseño de la evaluación fue el siguiente: proporcionar libros de texto a todos los estudiantes de un subconjunto de clases que originalmente estaban destinadas a servir de controles en un estudio en curso sobre la eficacia de programas de instrucción por radio. La mitad de las clases recibió libros de texto, la otra mitad no. Se aplicó una prueba preliminar a principios de año y una prueba final al término del año a todas las clases. Luego, en el estudio se aplicaron técnicas de regresión simples para comparar las puntuaciones medias de las pruebas finales de las clases como una función de las puntuaciones de las pruebas preliminares y la intervención.

Una importante lección que se aprendió es la forma de diseñar cuidadosamente una evaluación: la selección aleatoria mostró una construcción especialmente atinada y se combinó de manera inteligente con una prueba que permitió maximizar la comparabilidad cruzada. Se aprendió una segunda lección sobre el pragmatismo: en su diseño, la evaluación se anticipaba a potenciales aspectos de economía política bastante serios. Finalmente, la evaluación arroja una serie de ejemplos prácticos sobre el tipo de decisiones que se deben tomar durante el trabajo en terreno.

II. Preguntas de la investigación y diseño de la evaluación

El diseño de la evaluación contiene dos componentes interesantes: la concatenación con una evaluación preexistente y la comprensión preliminar del entorno político dentro del cual debía realizarse la evaluación. La pregunta clave de la investigación era directa: evaluar el impacto de una mayor disponibilidad de libros de texto en el aprendizaje de alumnos del primer grado, centrándose específicamente en la pregunta de si esos libros de texto efectivamente se usaban dentro de la sala de clases. Dado que ya existían una intervención a través de un programa de instrucción por radio (Matemáticas por Radio) la pregunta se amplió para comparar el impacto de la disponibilidad de libros de texto con la instrucción por radio, como asimismo con un grupo de control.

Cabe analizar la decisión de supervisar el uso efectivo de los libros de texto, lo que dificulta la evaluación. En muchas intervenciones educacionales se entregan materiales a las salas de clase, pero claramente el impacto de esos materiales depende de su uso. Sin

embargo, como lo destacan los evaluadores, esta decisión significa que la evaluación “no estima el potencial que los libros de texto o las lecciones radiales tienen para mejorar el rendimiento de los estudiantes bajo resultados óptimos. Más bien intenta evaluar su impacto en la forma en que se puedan adoptar en el típico país en desarrollo” (Jamison, 1981, página 559). Por este motivo, la simple provisión de libros de texto puede no ser suficiente por sí misma si no se diseña al mismo tiempo un método para asegurar que los maestros usen esos libros de texto en la forma prevista.

En la evaluación se utilizó un diseño aleatorio que se concatenó con la evaluación de un proyecto preexistente. En el Proyecto de Radio Nicaragua en curso, ya se habían implementado procedimientos completos de asignación y prueba aleatorios como mecanismo para evaluar la eficacia de un programa de instrucción por radio. En ese proyecto ya se habían clasificado todas las escuelas primarias en tres provincias de Nicaragua como grupo con instrucción de radio o de control, usando un proceso de muestreo aleatorio estratificado por urbanización (cerca del 30% de los estudiantes pertenecen a escuelas rurales, pero se seleccionaron números iguales de clases en cada estrato).

En la evaluación de los libros de texto se aprovechó este diseño existente seleccionando de la siguiente manera las escuelas de tratamiento y de control. En primer lugar, los evaluadores obtuvieron una lista de todas las escuelas con clases que cumplían los requisitos para cada una de las seis categorías (tres provincias, rurales y urbanas). A continuación, de estas listas maestras para cada categoría se asignaron aleatoriamente las escuelas a tratamiento o control y luego se usaron las escuelas en el orden en que aparecían (una escuela que se negó a participar fue sustituida por la siguiente en la lista). Se rechazaron las solicitudes de participar de las clases que pertenecían a los grupos de control y cualquier utilización de los materiales experimentales fue controlado por los autores. Cabe mencionar que el diseño de la evaluación había abordado desde su inicio esta potencial dificultad política. El equipo de evaluación anunció sus intenciones desde el inicio, obtuvo aprobación y apoyo oficial para la política y también estableció procedimientos claros y coherentes para el programa.

Por consiguiente, en el estudio se seleccionaron aleatoriamente 88 clases: 48 escuelas incluidas en el programa de radio y 40 de control. Veinte escuelas de control recibieron libros de texto para cada niño y los maestros recibieron tanto instrucciones por escrito como orales además de las versiones para los maestros de las pruebas. El componente radial constaba de 150 lecciones diarias de matemáticas, combinadas con hojas de trabajo para los estudiantes e instrucciones por escrito y orales para los maestros.

Un decisión inteligente que se adoptó fue la deliberada no supervisión de los grupos de tratamiento. Esto claramente era difícil, porque resultaba arduo evaluar la utilización del programa sin esta supervisión. Sin embargo, se juzgó que el costo de influir en el comportamiento era demasiado elevado. No se pudieron hacer visitas sorpresa, que fue la solución de compromiso aceptada, debido a conflictos políticos durante el año en que se realizó la evaluación; por tal motivo dichas visitas tuvieron que llevarse a cabo el año siguiente.

Una segunda decisión fue que las pruebas fueran administradas por el personal del proyecto y no por los maestros de las clases. Esto claramente elevó los costos administrativos, pero disminuyó un eventual sesgo al tomar las pruebas. Se hizo una prueba preliminar de la preparación para las matemáticas de los estudiantes durante las tres primeras semanas de asistencia a la escuela. Se tenía la intención de aplicar la prueba final, en que se medía el progreso alcanzado, durante las últimas tres semanas de escuela, pero fue realizada dos semanas antes debido a problemas políticos. Dentro de lo posible, los estudiantes tenían condiciones idénticas para las dos pruebas cuando fueron tomadas, porque tuvieron el mismo tiempo para realizar las pruebas y porque las instrucciones fueron grabadas en cinta.

III. Datos

Hay dos lecciones importantes que se pueden deducir del componente de recopilación de datos. La primera es que las dificultades logísticas a menudo son imposibles de evitar. A pesar del diseño cuidadoso, se presentaron una serie de problemas al momento de elaborar un conjunto perfecto de comparaciones de pruebas preliminares y finales. Si bien se tuvieron en total 20 clases de control, 20 clases con más libros de textos y 47 clases con lecciones por radio, los números de las puntuaciones de las pruebas preliminares y finales fueron diferentes en cada grupo debido a un registro tardío, deserción, ausentismo y la imposibilidad de administrar la prueba debido al exceso de alumnos. Aparentemente no se recopiló información individual sobre los estudiantes.

La segunda lección es el modo ingenioso en que los evaluadores diseñaron la prueba final para minimizar la carga y sin embargo obtener la información necesaria. Se tuvo que hacer frente a una serie de problemas:

- En Nicaragua no se utilizaban pruebas normalizadas.
- En la prueba se debía evaluar la consecución de los objetivos del programa de estudios.
- En la prueba se debía captar el aprovechamiento en cada materia, para facilitar una evaluación de la eficacia de la intervención para cada materia y general.

Para abordar estos aspectos, los evaluadores se basaron en un diseño de muestreo de matriz múltiple. La prueba incluía dos tipos de preguntas: las que se hacía a todos los alumnos de la clase (partidas 40 G) y las que se hacían a subconjuntos de estudiantes (partidas 44 I). Todas las partidas I se examinaron en todas las clases; la cuarta parte de todas las partidas G se examinó en cada clase. Esto permitió que los investigadores asignaran aleatoriamente unidades a lo largo de dos dimensiones: escuelas y formas de las pruebas. La puntuaciones medias de las pruebas finales para los grupos de tratamiento y grupos de control se derivaron sumando las puntuaciones promedio de cada prueba y los errores estándares se calcularon con la ayuda de la varianza residual después de eliminar los principales impactos de las partidas y los estudiantes.

El año después de la intervención también se recopiló información sobre la utilización de los libros de texto en 19 de las 20 escuelas en que se usaron libros de texto.

IV. Técnicas econométricas

La estructura de la evaluación significó que era adecuado realizar una comparación simple de las medias entre los grupos de tratamiento y de control, que fue lo que de hecho se utilizó. El método puede ser muy engorroso si existen múltiples estratos y múltiples intervenciones, que es lo que pasó en esta evaluación. Por consiguiente, los evaluadores también usaron un método de regresión simple. En este caso, la clase fue la unidad de análisis y se aplicó la regresión a la puntuación media de la prueba final de la clase con respecto a la puntuación media de la prueba preliminar y las variables ficticias para las intervenciones de radio y de libros de texto, una variable ficticia urbano/rural y la puntuación promedio de la prueba preliminar de la clase como variables independientes.

Un componente importante de cualquier evaluación es examinar si los diferentes grupos se ven afectados de manera diferente por el mismo tratamiento. A menudo es posible lograrlo, y así fue en esta evaluación, usando las variables interactivas de manera imaginativa. Se captaron las diferencias entre las áreas urbanas y rurales haciendo interactuar la variable ficticia urbano/rural con la intervención; se captó la diferencia en el impacto de la intervención sobre las puntuaciones de la prueba inicial haciendo interactuar las puntuaciones de la prueba inicial con la intervención.

V. Ejecutores

El Banco Mundial apoyó el proyecto de investigación, pero como parte de un proyecto conjunto de Matemáticas por Radio de la Agencia para el Desarrollo Internacional de los EE.UU. y el Ministerio de Educación de Nicaragua.

VI. Resultados

Los autores encontraron que tanto el tratamiento con libros de texto como a través de la radio tuvo impactos importantes en los resultados de los estudiantes: con la disponibilidad de libros de texto, las puntuaciones de los estudiantes en las pruebas finales aumentaron en 3,5 unidades correctas, las lecciones de radio en 14,9 puntos, lo que es notable si se considera que la desviación estándar de las clases es de 8,3 y de las unidades individuales de 11,8. Tanto las intervenciones con lecciones de radio como con libros de texto fueron más eficaces en las escuelas rurales y potencialmente podrían ser muy importantes para reducir la brecha entre la calidad urbana y rural de la educación. Estos resultados parecen ser independientes del nivel inicial de capacidades de la clase en término de las puntuaciones de las pruebas preliminares.

Los autores atribuyen la diferencia en los resultados de las intervenciones de radio y de libros de texto a diferencias en la utilización de los libros de texto, especialmente dado la deficiente educación de los maestros.

VII. Lecciones aprendidas

Se aprendieron tres lecciones principales: la importancia de las políticas en las decisiones de diseño; la utilidad de diseños de prueba imaginativos y las dificultades asociadas con el trabajo en terreno. En primer lugar, en este estudio se destacó la economía política de

diseño aleatorio: claramente existen presiones políticas bastante intensas que se podrían aplicar y que se deben abordar en una etapa temprana y con la ayuda del gobierno. En segundo lugar a través de una imaginativa aplicación del diseño de una prueba, los autores pudieron medir muchas facetas de los resultados del aprendizaje sin tener que realizar pruebas extensas poco realistas. Finalmente, los evaluadores claramente abordaron varios aspectos del trabajo en terreno: si supervisar o no la adopción efectiva de libros de texto y cómo hacerlo y quién debería administrar las pruebas.

VIII. Fuente

Jamison, Dean T., Barbara Serle, Klaus Galda y Stephen P. Heyneman. 1981 “*Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement*”. *Journal of Educational Psychology* 73(4): 556-67.

Anexo 1.13: El impacto de sistemas de recuperación de costos alternativos en el acceso y equidad en Níger

I. Introducción

Descripción del proyecto. La capacidad de recuperar una parte de los costos de la atención de salud es crucial en la prestación de servicios de salud. Sin embargo, es poco lo que se sabe sobre el impacto de diferentes estrategias en la calidad y los resultados en materia de bienestar. En la evaluación se estima el impacto en la demanda por atención de salud de dos sistemas piloto de recuperación de costos en el sector de atención primaria (no hospitalaria) en Níger. Níger es una economía rural pobre, los costos de la salud pública representan un 5% a 6% del presupuesto fiscal y una buena parte de este financiamiento está mal orientado hacia los hospitales y el personal. El gobierno deseaba evaluar las consecuencias de diferentes mecanismos de pago y consideró dos: un mecanismo de financiamiento que consta de un simple reembolso por atención prestada y otro de un tributo más reembolso por atención prestada, combinando ambos con mejoramientos en la calidad y la gestión. El gobierno estaba especialmente interesado en conocer la forma en que cambiaría la demanda por atención de salud, en especial entre los grupos vulnerables, y en examinar si tales mejoramientos de la calidad resultaban sostenibles.

Puntos destacados de la evaluación. Los diferentes mecanismos de pago se implementaron en tres distritos, uno para cada tratamiento y uno de control. En la evaluación se utilizó un diseño cuasi experimental basado en encuestas domiciliarias, combinadas con datos administrativos sobre la utilización y costos operativos. El especial interés de la evaluación radica en que aborda directamente los asuntos de economía política con un instrumento de encuesta que pregunta a los declarantes acerca de su disposición a pagar por el mejor servicio. Este reconocimiento explícito de que los resultados significativos no son por sí mismos suficientes para garantizar un proyecto sostenible constituye un aporte extremadamente valioso. Otro aspecto útil es la evaluación explícita del impacto de la intervención en diferentes grupos beneficiarios (niños, mujeres, aldeas sin un establecimiento de salud pública y los ciudadanos más pobres).

II. Preguntas de la investigación y diseño de la evaluación

Las principales preguntas se refirieron al impacto del tratamiento en (a) la demanda y utilización de los establecimientos de atención de salud públicos, (b) grupos beneficiarios específicos (pobres, mujeres y niños), (c) acceso financiero y geográfico, (d) el uso de servicios alternativos y (e) la sostenibilidad de los mejoramientos en el marco del mecanismo de recuperación de costos (costos de atención de pacientes y fármacos, como también ingresos y disposición a pagar).

De un registro administrativo se seleccionaron tres distritos de salud en distintas provincias. Si bien todos eran similares en términos de sus características económicas, demográficas y sociales, eran diferentes en términos étnicos. Cada distrito contaba con un centro médico, un centro de salud maternoinfantil, un puesto médico y un médico, como asimismo con dispensarios rurales.

En los dos distritos de tratamiento se implementaron cuatro mejoramientos en la calidad y la gestión; en el distrito de control no se implementó ninguno. Específicamente, se entregó una existencia inicial de medicamentos, el personal fue capacitado en diagnóstico y tratamiento, se instaló un sistema de manejo de existencias de medicamentos y de gestión financiera y se capacitó al personal en su utilización, además de aumentar la capacidad supervisora para reforzar la gestión.

Ambos mecanismos diferentes de tarificación se introdujeron simultáneamente. En el primero se trataba de una tarifa por episodio de 200 francos CFA (US \$ 0,66) por usuario de más de cinco años de edad y una tarifa de 100 francos CFA por usuario de menos de cinco años. En el segundo mecanismo se combinaba un tributo anual de 200 francos CFA pagado por los contribuyentes del distrito y una tarifa de 50 francos CFA por usuario de más de cinco años y 25 francos CFA por niños menores de cinco. El ingreso anual era inferior a \$300 per cápita. En cada sistema se incluían exenciones para grupos beneficiarios específicos. Los fondos se administraron a nivel de distrito.

III. Datos

Los tres distritos se seleccionaron a partir de datos administrativos. Se implementaron dos encuestas domiciliarias, una básica o de referencia, y éstas se combinaron con los registros administrativos sobre los establecimientos. En cada encuesta se recopiló información demográfica sobre los hogares e individuos de una muestra de 1.800 hogares seleccionados aleatoriamente. La encuesta básica contaba con información sobre 2.833 individuos que habían estado enfermos las dos semanas anteriores a la encuesta y sobre 1.770 mujeres embarazadas; la encuesta final contó con información sobre 2.710 individuos enfermos y 1.615 mujeres embarazadas. En los datos administrativos había información bastante detallada sobre los gastos mensuales en consumo de fármacos y administración, mantenimiento de personal y recepción de honorarios, junto con la utilización de los establecimientos de salud. Esta información se recopiló el año anterior a la intervención, el año básico (mayo 1992 a abril 1993) y el año posterior a la intervención.

IV. Técnicas econométricas

En el estudio se combinan las comparaciones de medias con técnicas logit simples, utilizándose estas últimas para captar los cambios en el patrón de utilización; específicamente, la respuesta individual de si se usó o no el establecimientos de atención de salud (P_1), para especificar el siguiente modelo:

$$\text{Logit}(P_1) = X + * (A+B)$$

Este modelo, que controla por un vector de las características individuales, X, y las variables ficticias A y B, se comparó con

$$\text{Logit}(P_i) = X + \alpha A + \beta B$$

Las variables ficticias A y B se definen de diversas maneras. En la primera serie de regresiones, A se refiere al período durante el tratamiento, B al período antes del tratamiento y las regresiones se ejecutan por subgrupo (los grupos beneficiarios especificados) y por distrito. En la segunda serie de regresiones, A y B se utilizan para ejecutar 6 comparaciones por pares de cada uno de los distritos con cada uno de los demás distritos durante el tratamiento. En cada caso, los autores comprueban si $(\alpha + \beta) = \alpha$. Los impactos del acceso geográfico y financiero se captan en la matriz X mediante medidas de la distancia caminada a pie y los cuartiles de ingreso, respectivamente. Del análisis no se deduce claramente cuál es la categoría omitida en cada caso. Tampoco queda claro si los errores estándares de las estimaciones se corrigieron por la naturaleza combinada del diseño de la muestra.

Si bien las técnicas logit son un modo eficiente para abordar tres de las cuatro preguntas de la investigación (patrones de utilización, el impacto en subgrupos y los impactos del acceso geográfico y financiero), la cuarta pregunta, el impacto de los cambios en la recuperación de costos, se aborda a través de los datos administrativos y comparaciones simples de las medias. Una preocupación obvia en este último enfoque, que no se abordó explícitamente, es la posibilidad de sesgo en la información de los resultados postratamiento. Específicamente, existe un cierto riesgo moral cuando los administradores se evalúan respecto a la respuesta exitosa al tratamiento.

El impacto de los tratamientos en el uso de sistemas de salud alternativos se abordó a través de técnicas econométricas que se describen en otra parte.

V. Ejecutores

La encuesta fue realizada por el Ministerio de Salud Pública, con la asistencia financiera y técnica de la USAID y el Banco Mundial. La evaluación misma fue realizada por Francis Dip, Abode Yazbeck y Ricardo Bitran, de Abt Associates.

VI. Resultados

En el estudio se determinó que el tributo más reembolso generaba más ingresos per cápita que el sistema basado sólo en reembolso, además de ser mucho más popular. El primer sistema también arrojó mejores resultados en materia de prestar acceso a una mejor atención de salud a los pobres, mujeres y niños. Sin embargo, dado que el aspecto geográfico representa una de las principales barreras al acceso a la atención de salud, un sistema basado en tributo de hecho redistribuye el costo de la atención de salud de las personas que viven cerca de los establecimientos de salud a las personas que viven alejadas de éstos.

En los distritos donde se implementó el sistema de reembolso por atención prestada, el número de visitas iniciales disminuyó levemente, pero aumentó la demanda por servicios de atención de salud, a diferencia del drástico aumento de ambos en el distrito donde se implementó el sistema de tributo más reembolso. Gran parte de ello es atribuible a la mejor calidad del servicio asociada con los mejoramientos de calidad, lo que más que compensa el aumento de costo.

El control de los costos asociado con la reforma en materia de calidad y gestión, especialmente los costos de fármacos, también demostró ser eficaz y sostenible. La recuperación de costos en el distrito donde se implementó el tributo más reembolso alcanzó e incluso superó el 100%, pero fue sustancialmente menor en el distrito donde se implementó el reembolso por servicio prestado. Más aún, en el primero hubo mucho más disposición a pagar que en el segundo.

El principal resultado es que el método de tributo más reembolso es mucho más eficaz en lograr las metas propuestas y cuenta con mayores simpatías de la población. Sin embargo, en la evaluación también se demostró que la falta de acceso geográfico a los establecimientos de atención de salud es una importante barrera a la utilización. Esto apunta a que existen algunos aspectos de distribución asociados a la transición a un sistema de tributo más reembolso: los hogares que están muy alejados de los establecimientos de atención de salud implícitamente estarían subvencionando a los hogares ubicados cerca.

VII. Lecciones aprendidas

En esta evaluación hay varias lecciones útiles. Una es la forma multifacética en que se evalúa el impacto del proyecto sobre múltiples dimensiones relacionadas con la sostenibilidad: no sólo en la recuperación de costos, sino también en la calidad y la reacción de los grupos beneficiarios afectados. Otra es la atención al detalle en la recopilación de datos con instrumentos administrativos y de encuestas, lo que rindió sus frutos a través de la capacidad para identificar exactamente los componentes de la intervención que funcionaron y su motivo. Finalmente, el análisis del impacto en cada grupo beneficiario resultó ser especialmente útil para formular recomendaciones de políticas.

VIII. Fuentes

Diop, F. A. Yazbeck, y R. Bitran. 1995. “*The Impact of Alternative Cost Recovery Schemes on Access and Equity in Niger*”. *Health Policy and Planning* 10(3): 223-40.

Wouters, A. 1995. “*Improving Quality through Cost Recovery in Niger*”. 10(3): 257-70.

Anexo 1.14: Resultados de la instrucción en las escuelas primarias de Filipinas: Evaluación del impacto de cuatro experimentos

I. Introducción

Descripción del proyecto. En la mayoría de los países en desarrollo, las elevadas tasas de deserción y el inadecuado nivel de aprendizaje de los alumnos en la educación primaria preocupan a los gestores de políticas. Esto ciertamente es el caso de Filipinas: casi la cuarta parte de los niños filipinos se retiran antes de completar el sexto grado y aquellos que terminan a menudo sólo dominan la mitad de lo que les enseñaron. El gobierno emprendió un Programa de Intervención contra la Deserción (DIP, *Dropout Intervention Program*) en 1990-92 para enfrentar estos problemas. Se realizaron cuatro experimentos: suministro de materiales de aprendizaje multinivel (MLM, *multi-level learning materials*), almuerzos escolares (SL, *school lunches*) y una combinación de cada uno de éstos con una colaboración entre padres y maestros (PTP, *parent-teacher partnership*). El primer método permite que los maestros adapten el ritmo de la enseñanza a las diferentes necesidades de los estudiantes y es mucho menos costoso que la alimentación escolar. Las colaboraciones entre padres y maestros no cuestan casi nada, pero pueden ayudar a los estudiantes a aprender tanto en el hogar como en la escuela.

Puntos destacados de la evaluación. Esta evaluación es digna de destacar por su orientación explícita a la construcción de capacidades en el país de origen, de modo que la evaluación pueda transformarse en un componente integral de nuevas iniciativas y que los requisitos respecto a datos se considerarán más bien antes de la implementación de nuevos proyectos que después. Sin embargo, a consecuencia de ello surgen algunos problemas y la evaluación es muy clara acerca de lo que se puede esperar. Otro aporte importante de la evaluación es que verifica la solidez de los resultados con diferentes métodos econométricos. Finalmente, el análisis de costos y beneficios que se aplicó al final es importante en que explícitamente reconoce que no basta con obtener resultados significativos: las intervenciones poco costosas pueden ser incluso mejores que aquellas que son onerosas.

II. Preguntas de la investigación y diseño de la evaluación

El aspecto clave de la investigación es evaluar el impacto de cuatro intervenciones diferentes en la deserción escolar y los resultados de los estudiantes. Sin embargo, el diseño de la evaluación está condicionado tanto por necesidades pragmáticas como programáticas. El equipo del programa DIP siguió un proceso de selección de escuelas de tres etapas:

- Se identificaron dos distritos en cada una de las cinco regiones del país como municipios de bajo ingreso. En un distrito, las opciones de tratamiento se integraron en un paquete de control, MLM o MLM-PTP; en el otro, de control, SF o SF-PTP. La asignación de los dos paquetes de intervenciones fue aleatoria.

- En cada distrito, el equipo seleccionó tres escuelas que (a) tenían todos los grados de instrucción, con una clase por grado, (b) tenían altas tasas de deserción y (c) no contaba con programa de alimentación escolar.
- Las tres escuelas en cada distrito se asignaron ya sea a control o a una de las dos intervenciones, sobre la base de un sorteo aleatorio.

Cada intervención se asignó aleatoriamente a todas las clases en cinco escuelas y en 1991 y en 1992 se aplicaron tanto pruebas preliminares como finales a todas las clases en las 20 escuelas, como asimismo en las 10 escuelas de control.

III. Datos

El procedimiento de recopilación de datos es instructivo por sí sólo. La recopilación de los datos básicos se inició en 1990-91 y las intervenciones se implementaron en 1991-92. En cada uno de los dos años se reunió información detallada sobre 29 escuelas, aproximadamente 180 maestros y 4.000 alumnos. Aunque estos cuestionarios fueron muy detallados, resultó que esto fue innecesario: en realidad sólo se utilizó un pequeño subconjunto de información, lo que parece indicar que parte de la carga del proceso de evaluación se podría minimizar provechosamente. También se aplicaron pruebas preliminares y finales en tres materias al principio y término de cada año escolar: matemáticas, filipino e inglés.

La información se estructuró para que fuera longitudinal tanto sobre los alumnos como sobre las escuelas. Desafortunadamente, los identificadores de los estudiantes resultaron no ser exclusivos para los alumnos y escuelas entre los dos años. Cabe señalar que esto no se sabía *a priori* y sólo se hizo evidente una vez que seis meses de trabajo revelaran inconsistencias internas. No fue posible recuperar los identificadores originales del Departamento de Educación filipino. La información afortunadamente se pudo rescatar para los alumnos de primer grado, permitiendo realizar algo de análisis longitudinal.

IV. Técnicas econométricas

La estructura del procedimiento de muestreo planteó algunos problemas econométricos interesantes: un conjunto para los desertores y un conjunto para los resultados de las puntuaciones en las pruebas. En cada caso se presentan dos conjuntos de controles obvios: uno es el grupo de control de las escuelas y el otro es la encuesta básica realizada el año anterior a la intervención. Los autores los manejaron de modos diferentes.

En el análisis de la deserción, es lógico establecer un método de diferencia en las diferencias y comparar el cambio en la tasa de deserción media en cada clase de la intervención entre los dos años con el cambio en la tasa de deserción para las clases de control. Sin embargo, inmediatamente surgieron dos problemas. En primer lugar, si bien los resultados fueron bastante considerables en magnitud, sólo fueron significativos para la intervención de MLM, posiblemente debido al problemas del pequeño tamaño de la muestra. Esto no es raro en este tipo de procedimientos y es probable que sea endémico debido a la falta de financiamiento para experimentos a gran escala en el contexto de un país en desarrollo. En segundo lugar, una breve verificación de si las características y

resultados de los estudiante en realidad eran los mismos en todas las escuelas en el año anterior a la intervención parece indicar que habían algunas diferencias importantes en las características. Estos dos factores llevaron a los autores a examinar la solidez de los resultados a través de técnicas de regresión logística que controlaban por las características personales (PC, *personal characteristics*) y el trasfondo familiar (FB, *family background*). El resultado central no cambió. Sin embargo, la técnica de regresión sí develó una importante causa central indirecta de la deserción, que fue el deficiente desempeño académico. Esto naturalmente llevó a un segundo conjunto de análisis, el que se centró en el rendimiento.

Se planteó un conjunto distinto de inquietudes econométricas al evaluar el impacto de la intervención INTER en el desempeño académico del individuo I en la escuela s en el momento t (AP_{ist}), que los autores modelan como:

$$AP_{ist} = \delta_0 + \delta_1 AP_{ist-1} + \delta_2 PC_i + \delta_3 FB_i + \delta_4 LE_{st} + \delta_5 CC_i + \delta_6 INTER_{jt} + \varepsilon,$$

donde LE es el entorno de aprendizaje y CC son las condiciones de la clase.

Dentro de éstos, el problema principal es dar cuenta de la probable correlación combinada de errores para estudiantes en las mismas clases y escuelas. El segundo es el intento de captar la heterogeneidad no observada. Y el tercer problema, aunque relacionado, es el sesgo de selección.

Para manejar el primer problema, a los errores estándares se aplica una corrección Huber-White. En principio, el segundo se podría captar a nivel individual usando la diferencia en las puntuaciones de las pruebas como una variable independiente. Sin embargo, los autores sostienen que esto no es adecuado, ya que presupone que el valor de δ_1 es 1, lo que no es validado por pruebas. Por este motivo, ellos mantienen la especificación de la variable dependiente desfasada, lo que sin embargo plantea el siguiente problema, uno de sesgo del regresor endógeno. Para manejarlo, se instrumentaliza la puntuación de las pruebas preliminares en cada sujeto con las puntuaciones de las pruebas preliminares en los otros sujetos. No obstante, los autores hacen notar que la reducción en el sesgo conlleva un costo, una menor eficiencia, y por consiguiente informan resultados de variables instrumentales y mínimos cuadrados. Los autores utilizan impactos fijos a nivel de escuela y maestros para controlar por la heterogeneidad no observada en el entorno de aprendizaje (LE, *learning environment*) y las condiciones pedagógicas (CC, *classroom conditions*).

El tercer problema es uno que también es endémico a la literatura y para el cual no existe una solución plenamente aceptada: el sesgo de selección. Claramente, dado que existen tasas de deserción diferenciales, el rendimiento académico individual está condicionado a la decisión de no desertar. Si bien este problema a menudo se abordó con el procedimiento de Heckman de dos etapas, por los siguientes tres motivos hay bastante disconformidad con él: su sensibilidad al supuesto de la distribución normal, la elección y conveniencia de la variables adecuadas que hay que usar en la primera etapa y su frecuente dependencia de la identificación a través de la no linealidad de la primera etapa.

Desafortunadamente, todavía no hay consenso en torno a una alternativa adecuada. Una de ellas fue propuesta por Krueger, quien asigna las calificaciones de las pruebas preliminares a las deserciones y las retorna a la regresión. Por consiguiente, los autores informan tres conjuntos de resultados: la simple regresión de los resultados en contra de la intervención; el enfoque Krueger y el procedimiento Heckman.

V. Ejecutores

La recopilación de datos fue realizada por la Oficina de Educación Primaria del Departamento de Educación, Cultura y Deportes de Filipinas. El análisis fue ejecutado por un funcionario del Banco Mundial y dos investigadores académicos.

VI. Resultados

En el estudio se evalúa el impacto de estas intervenciones sobre la deserción en los grados uno hasta seis y sobre los resultados en las puntuaciones de las pruebas en el primer grado usando el método de diferencia en las diferencias, técnicas de variables instrumentales y el método de selección de Heckman. El impacto de los materiales multinivel sobre la deserción y el mejoramiento del rendimiento académico (especialmente en combinación con una colaboración entre padres y apoderados) es sólido a diferentes especificaciones, además de ser bastante eficaz en función de los costos. El impacto de los almuerzos escolares en general fue débil. Un componente interesante del estudio fue un análisis de costos-beneficios, que establece la importante proposición ¡de que esto no arroja resultados significativos! Específicamente, al hacer un cálculo simple de los costos (de oportunidad) directos e indirectos del programa, se llega a la conclusión que el enfoque MLM es tanto efectivo como eficaz en función de los costos.

La falta de eficacia de la alimentación escolar quizás esté sobredimensionada: posiblemente un método más focalizado para los programas de alimentación escolar resultaría adecuado. Además, dado que el tiempo transcurrido entre la implementación y la evaluación del programa es bastante corto, en la evaluación no se puede abordar el impacto a largo plazo de las intervenciones.

VI. Lecciones aprendidas

Este procedimiento de evaluación permitió aprender varias lecciones. Una de las principales es lo diabólico de los detalles, esto es, que se puede perder una gran cantidad de información longitudinal vital si se pierde la información adecuada, como la unicidad de los identificadores a través del tiempo. Una segunda es que se utilizó muy poco de la información que se recopiló en detalladas encuestas y que se podría haber disminuido una parte sustancial de la carga que se impuso a los declarantes. Tercero, en el estudio se destaca el valor de diferentes métodos econométricos y las ventajas de encontrar coherencia entre las técnicas. Cuarto, este estudio es ejemplar en la forma en que utiliza el análisis de costos-beneficios, al identificar y también valorar los costos de las diferentes intervenciones. Y finalmente, aunque claramente se incurrió en errores durante el estudio, los autores advierten que uno de los motivos más importantes para el estudio fue la construcción de capacidades de evaluación dentro del país. El hecho que se haya

implementado y evaluado el Programa de Intervención de la Deserción significa que es posible crear dicha capacidad dentro de los ministerios de educación.

VII. Fuente

Tan, J.P, J. Lane y G. Lassibille. 1999. "Schooling Outcomes in Philippine Elementary Schools: Evaluación of the Impact of Four Experiments". *World Bank Economic Review*, septiembre.

Anexo 1.15: Evaluación del impacto de los proyectos de caminos rurales en la pobreza en Vietnam

I. Introducción

Descripción del proyecto. El Banco Mundial y otros donantes están promoviendo vigorosamente la construcción de caminos rurales como instrumentos para aliviar la pobreza. El Proyecto I de transporte rural de Vietnam se inició en 1997 con financiamiento del Banco Mundial para ser implementado en el transcurso de tres a cinco años. El objetivo del proyecto es elevar las condiciones de vida en áreas pobres mediante la rehabilitación de caminos y puentes existentes y el mejoramiento del acceso a los mercados. En cada provincia participante, para la identificación de los proyectos de rehabilitación se aplican los criterios de costo mínimo (tamaño de la población que se beneficiará y costo del proyecto). Sin embargo, en un esfuerzo por mejorar la orientación hacia el objetivo de disminuir la pobreza, un 20% de los fondos de cada provincia se puede destinar a áreas montañosas de baja densidad habitadas por minorías étnicas, donde los proyectos estrictamente no calificarían en el marco de los criterios de costo mínimo.

Evaluación del impacto. Pese al consenso generalizado acerca de la importancia de los caminos rurales, la evidencia concreta acerca del tamaño y la naturaleza de los beneficios de esa infraestructura sorprende por lo escasa. Con la Evaluación del impacto de los caminos rurales en Vietnam se busca determinar de qué manera cambia el bienestar de los hogares en las comunas que cuentan con intervenciones del proyecto vial si se comparan con otras que no las tienen. El problema clave de la evaluación es lograr aislar efectivamente el impacto vial de otros miles de factores que actualmente están cambiando en Vietnam rural, como resultado de la presente transición a una economía de mercado.

La evaluación se inició en forma paralela a la preparación del proyecto a principios de 1997 y está en marcha. Todavía no se dispone de resultados. La evaluación es urgente, ya que se trata de uno de los primeros intentos globales por evaluar el impacto de un proyecto vial rural en materia de bienestar: la última palabra en términos de evaluar si los proyectos realmente reducen la pobreza. Se pretende mejorar su diseño en relación con anteriores esfuerzos de evaluación de la infraestructura, combinando los siguientes elementos: (a) recopilación de datos de encuestas básicas y de seguimiento, (b) inclusión de controles adecuados, de modo que los resultados sean resistentes a factores no observados que influyen tanto en la ubicación del programa como en los resultados y (c) el seguimiento del proyecto durante un tiempo suficiente (mediante sucesivas rondas de recopilación de datos) para poder captar su pleno impacto en el bienestar.

II. Diseño de la evaluación

El diseño de la Evaluación del impacto de los caminos rurales de Vietnam se centra en los datos de encuestas básicas (antes de la intervención) y de seguimiento (después de la intervención) para una muestra de comunas con y sin proyecto. A través de técnicas de comparación pareada, se pueden identificar controles adecuados del universo de las

comunidades sin proyecto. Los datos básicos permiten realizar una comparación (“reflexiva”) antes y después de los indicadores del bienestar en la comunidades del grupo con proyecto y de control. En teoría, el grupo de control, que se selecciona mediante técnicas de comparación pareada, es idéntico al grupo del proyecto tanto en términos de características observadas como no observadas, de modo que los resultados que surjan en las comunidades del programa se pueden atribuir a la intervención del proyecto.

III. Técnicas de recopilación y análisis de datos

La información recopilada para fines de la evaluación incluyen encuestas a nivel comunal y de hogar, junto con bases de datos a nivel de distrito, provincias y proyecto. Las encuestas básicas y de seguimiento de comunas y hogares se realizaron en 1997 y 1999, y están programadas una tercera y cuarta ronda de encuestas para ser realizadas con dos años de intervalo. La muestra de las encuestas incluye 100 comunas con proyecto y 100 comunas sin proyecto, ubicadas en 6 de las 18 provincias que abarca el proyecto. Las comunas con proyecto se seleccionaron aleatoriamente de las listas de todas las comunas con propuesta de proyectos en cada provincia.. A continuación se redactó una lista de todas las demás comunas en los distritos donde se propusieron proyectos, de las cuales se escogieron aleatoriamente las comunas de control (idealmente, los grupos de control difieren del grupo con proyecto exclusivamente en que no contaron con una intervención. Además, por motivos logísticos era aconsejable limitar el trabajo en terreno a determinadas regiones. Por consiguiente, los controles se escogieron de la vecindad, y obviamente de los mismos distritos, que las comunas con tratamiento. Los distritos son grandes y por ende no es probable que se haya producido una contaminación de la comuna con proyecto a la comuna sin proyecto, pero este aspecto deberá verificarse cuidadosamente). Para probar la selección de los controles, se utilizarán técnicas de correspondencia de puntuación de la propensión basadas en las características de las comunas y todas las comunas de control que tengan atributos poco usuales en relación con la comunas del proyecto se eliminarán de la muestra. Se estimará y usará un modelo logit de la participación de las comunas en el proyecto, para asegurar que las comunas que se utilicen como control tengan puntuaciones de propensión similares (valores predichos del modelo logit).

La base de datos de las comunas se basa en los datos administrativos existentes recopilados anualmente por las comunas y cubren los aspectos demográficos, uso de la tierra y actividades de producción, incrementada con los datos de una encuesta a nivel de comuna realizada para los fines de la evaluación. La encuesta cubre las características generales, infraestructura, empleo, fuentes de sustento, agricultura, tierra y otros activos, educación, atención de salud, programas de desarrollo, organizaciones de la comunidad, finanzas comunales y precios. Estos datos se utilizarán para construir varios indicadores comunales del bienestar y para comprobar los impactos del programa en el tiempo.

El principal objetivo de la encuesta domiciliaria es captar información sobre el acceso de los hogares a varios recursos y servicios y los cambios de éstos en el tiempo. El cuestionario domiciliario se aplicó a 15 hogares seleccionados aleatoriamente en cada

comuna y cubrió aspectos de empleo, bienes, actividades productivas y laborales, educación, salud, comercialización, crédito, actividades comunitarias, acceso a programas de seguridad social y contra la pobreza, y transporte. Debido a la restringida capacidad dentro del país para realizar encuestas, no se intentó recopilar el conjunto completo de datos necesario para generar un indicador del bienestar a nivel de hogar (como ingreso o consumo). Sin embargo, en la encuesta se incluyeron varias preguntas que eran una réplica de las preguntas de la Encuesta de Niveles de Vida de Vietnam (ENV). Con la ayuda de ésta y otra información sobre las características de los hogares que son comunes en ambas encuestas, se usarán técnicas de regresión para estimar la situación de cada hogar dentro de la distribución nacional del bienestar. También se preparó una pequeña base de datos a nivel de distrito para ayudar a posicionar los datos a nivel de la comuna dentro del contexto, entre ellos los datos sobre población, uso de la tierra, economía e indicadores sociales. Cada una de estas encuestas se deberá repetir de acuerdo al programa de encuestas comunales.

Con la información disponible se establecieron dos bases de datos adicionales. Una, muy amplia a nivel provincial, para ayudar a entender la selección de las provincias dentro del proyecto. Esta base de datos cubre todas las provincias de Vietnam y contiene información sobre un gran número de variables socioeconómicas. Y la otra, una base de datos a nivel provincial para cada una de las áreas encuestadas de proyecto, para controlar por la magnitud del proyecto y su método para implementar la evaluación de impacto del proyecto.

Los datos básicos se utilizarán para modelar la selección de los sitios de aplicación del proyecto, centrándose en los procesos económicos, sociales y de economía política subyacentes. Se utilizarán rondas posteriores para entender las ganancias que se pueden medir a nivel de comuna, condicionadas a la selección. El enfoque analítico serán los métodos de “doble diferencia” con correspondencia. Se utilizará la correspondencia para seleccionar los controles ideales de entre 100 comunas de la muestra no incluidas en el proyecto. Se compararán los resultados en las comunas donde se aplicó el proyecto con la información que se encuentre en las comunas de control, tanto antes como después de la incorporación de los proyectos de caminos. A continuación se identifica el impacto del programa como la diferencia entre los resultados de las áreas con proyecto después y antes del programa, menos las diferencias correspondientes en los resultados en las áreas de control pareadas. Esta metodología arroja una estimación no sesgada de los impactos de un proyecto en presencia de factores no observados invariables en el tiempo que influyen tanto en la selección de las áreas de los proyectos como en los resultados. Los resultados se verán perfeccionados por la riqueza de los conjuntos de datos en indicadores de resultados y variables explicativas. Los indicadores de resultados que se examinarán incluyen rendimientos agrícolas a nivel comunal, diversificación de las fuentes de ingreso, oportunidades de empleo, uso y distribución de la tierra, disponibilidad de bienes, servicios y recursos y la profusión y distribución de activos.

IV. Costos y administración de la evaluación

Costos. Hasta la fecha, el costo total de la evaluación es de \$222.500 o el 3,6% del costo total del proyecto. Esta suma incluye \$202.500 para cubrir las dos primeras rondas de recopilación de datos y una donación de \$20.000 para investigación. En estos costos no se incluyen los costos de tiempo y gastos de viaje del personal del Banco Mundial.

Administración. La evaluación fue diseñada por Dominique van de Walle, miembro del equipo del Banco Mundial. Como supervisor del estudio dentro del país se contrató a un consultor independiente con experiencia en economía e investigación en pobreza rural y desarrollo. Este consultor contrató y capacitó a los supervisores de los equipos, organizó todos los aspectos logísticos y supervisó toda la recopilación de datos.

V. Fuente

Van de Walle, Dominique. 1999. *Assessing the Poverty Impact of Rural Road Projects*. Banco Mundial. Washington, D.C. Procesado.

Anexo 2: Muestra de términos de referencia

Ejemplo I. Proyecto de nutrición y desarrollo de la primera infancia de Uganda

Términos de referencia para la asesoría de la empresa consultora en la evaluación del proyecto

I. Antecedentes

El Gobierno de Uganda (GOU) ha solicitado un préstamo de la Asociación Internacional de Fomento (AIF) para financiar el costo de un Proyecto de Nutrición y Primera Infancia. El proyecto se centra en mejorar la calidad de vida de niños menores de seis años y crear la capacidad de las familias y las comunidades para cuidar a sus niños. Específicamente, con el proyecto se apunta a lograr el desarrollo de la primera infancia mejorando aspectos como la nutrición, salud y situación sicosocial y cognitiva de los niños menores de seis años en Uganda.

I. Fundamentos para invertir en el desarrollo de la primera infancia

La inversión en el desarrollo de la primera infancia tiene beneficios concretos no sólo para los niños y sus padres, sino que para la comunidad y el país en su conjunto. Durante los primeros años de vida y la preescolar se produce un rápido crecimiento físico y desarrollo mental; a los dos años de vida, el cerebro de un niño ha crecido casi hasta su máximo definitivo. Las capacidades cognitivas también se desarrollan en gran medida hasta los cuatro años de edad. El adecuado crecimiento y desarrollo físico y mental durante la primera infancia mejoran la preparación para la escuela, acrecientan la retención escolar y contribuyen a la dependencia en el capital humano. Los niños provenientes de ambientes menos aventajados se pueden beneficiar especialmente de una atención durante su primera infancia, acortando con ello las brechas e inequidades relacionadas con la pobreza.

Una buena salud y nutrición son aspectos clave, al igual que la estimulación mental, si el niño ha de desarrollar estructuras conceptuales sólidas más adelante en su vida. La sinergia entre nutrición, salud y estimulación mental es de tal importancia que los efectos positivos concretos en el crecimiento y desarrollo del niño sólo se pueden lograr a través de un enfoque integrado.

I. Objetivos y estrategias del proyecto

En términos de desarrollo, el objetivo del proyecto es mejorar el crecimiento y desarrollo de los niños menores de seis años en materia de nutrición, salud y aspectos sicosociales y cognitivos. El logro de estos objetivos al final del período de implementación de cinco años se medirá de acuerdo con los siguientes marcadores: (a) reducción de niños preescolares de bajo peso en un tercio con respecto a los niveles de 1995 en los distritos donde se ejecutó el proyecto, (b) reducción del retardo en el crecimiento al ingresar a las escuelas primarias de un 25% con respecto a los niveles de 1995 en los distritos donde se ejecutó el proyecto, (c) mejoramiento en el desarrollo sicosocial y cognitivo de los niños, (d) menores tasas de repetición y deserción escolar en el nivel inferior de la educación primaria y (e) desarrollo de capacidades empresariales y habilitación económica de las madres y encargados del cuidado de los niños.

El proyecto presta su apoyo al Programa Nacional de Uganda de Acción para los Niños y el Plan de Acción para la Erradicación de la Pobreza. En este proyecto se da especial realce al mejoramiento de la preparación de los niños menores para el colegio, con lo cual se contribuye al logro de una educación primaria universal. La principal estrategia del proyecto es mejorar la capacidad de las familias y las comunidades para cuidar mejor de los niños en edad preescolar (de cero a seis años) al mejorar el conocimiento sobre el crecimiento y desarrollo infantil, la función de los progenitores, la nutrición y el cuidado de la salud y actividades generadoras de ingresos para la mujeres.

I. Estrategia del proyecto

El proyecto es un programa con prioridades locales determinado por los procesos mismos, más que un paquete detallado, planificado con anterioridad. Los aportes se introducirán gradualmente en las comunidades como resultado de un proceso participativo para asegurar el sentido de pertenencia y la sostenibilidad. El programa implicará la colaboración entre entidades gubernamentales y no gubernamentales, inclusive ONG locales e internacionales, y las comunidades. Como programa multisectorial que involucra componentes de salud, nutrición, educación de la primera infancia, cuidado de niños, ahorros y generación de ingresos, el enfoque implicará la vinculación de diferentes departamentos gubernamentales y entidades no gubernamentales para proporcionar un servicio integral orientado al desarrollo de los niños. Con el proyecto se apoyarán una gama de opciones (un menú de programas) relacionadas con las necesidades de los niños preescolares y sus familias.

I. Componentes del proyecto

Componente 1 del proyecto– Intervenciones comunitarias integradas para el cuidado de los niños. Con este componente se pretende apoyar las metas del Gobierno tendientes (a) a mejorar la conciencia de los padres y madres respecto a los principales aspectos del cuidado, crecimiento y desarrollo de los niños a través de la educación de los padres, la supervisión del crecimiento de los niños y la promoción, capacitación y sensibilización; y (b) habilitar a las comunidades para que apoyen los programas de desarrollo infantil a través de la creación de capacidades, habilidades para la generación de ingresos y donaciones de apoyo. El objetivo es reducir en un tercio la desnutrición (bajo peso para la edad) infantil al final del período de cinco años en aquellos distritos donde se lleve a cabo el proyecto y mejorar la preparación de los niños para la escuela primaria, contribuyendo con ello al objetivo de contar con una educación primaria universal. En último término, el plan de Gobierno es abarcar todos los distritos; sin embargo, las intervenciones en esta fase se ejecutarán en 25 distritos escogidos por el Gobierno en base al nivel de desnutrición, mortalidad infantil y tasa de matrícula en la educación primaria. En el proyecto se incluyen las siguientes intervenciones relacionadas entre sí:

(a) Educación de padres y madres. Con este subcomponente se acrecentará el entendimiento de los progenitores y aquellos a cargo del cuidado de los niños sobre los principales aspectos de su crecimiento y desarrollo, inclusive nutrición y salud infantil y desarrollo sicosocial y cognitivo. En los padres se fortalecerá una amplia gama de competencias relacionadas. La creación de capacidades y conocimientos en los progenitores a su vez redundará en una mejor salud, desarrollo sicosocial y bienestar de los niños y en última instancia, su receptividad a la educación a nivel primario. En el marco del programa se movilizarán grupos de madres (y padres) a nivel comunitario, apoyados por materiales del proyecto en sus idiomas locales, supervisión técnica y comunicaciones. Se han probado con éxito materiales de aprendizaje simplificado para adultos con bajo nivel de alfabetismo en Uganda. El énfasis se pondrá en el mejoramiento de prácticas de cuidado infantil que promuevan el crecimiento y desarrollo adecuado de los niños, inclusive nutrición y salud infantil (como prácticas de amamantamiento y destete, especialmente el período de introducción de alimentos sólidos, al igual que el tipo de alimentos que se proporcionen y su preparación, promoción del crecimiento infantil y desparasitamiento), desarrollo sicosocial, estimulación cognitiva y apoyo social, además de higiene y mejores prácticas de salud en el hogar.

Estas intervenciones se verán fortalecidas y contarán con el apoyo de una actividad de extensión (el día de la infancia) organizada a nivel de subdivisión de condados para permitir que las comunidades accedan a varios servicios relacionados con la infancia en un solo lugar. También se llevará a cabo un estudio del impacto que tiene la entrega de *anathelminth albendazole* a niños pequeños en subdivisiones

seleccionadas de condados en el transcurso de los días de la infancia en cada subdivisión, para medir el efecto de tratamientos semestrales en el aumento del peso.

(b) Creación de capacidades y habilitación de la comunidad para el cuidado de los niños. Este subcomponente comprende dos actividades relacionadas: (a) creación de capacidades en las comunidades a través de talleres de sensibilización y planificación comunitaria (b) capacitación en habilidades empresariales para aumentar los ingresos de las madres y de aquellos a cargo del cuidado de los niños.

Componente 2 del proyecto– Donaciones de apoyo a las comunidades para el desarrollo de los niños. Hay dos tipos de donaciones disponibles para las comunidades:

(a) Donaciones de apoyo a comunidades – donaciones a las comunidades que se ofrecen sobre la base de equiparar las contribuciones de las propias comunidades. Estas donaciones y contribuciones de las comunidades cubrirán actividades diseñadas para apoyar intervenciones orientadas al desarrollo infantil que caben dentro de las directrices y menús contenidos en el manual de implementación del proyecto. Para cumplir los requisitos y obtener esta donación, las comunidades proporcionan aportes de contrapartida, las que pueden tomar la forma de bienes, trabajo o servicios. Entre los ejemplos de los usos de dichas donaciones está la construcción y operación de guarderías infantiles comunitarias, guarderías infantiles en el hogar o la producción y comercialización de alimentos para el destete. El componente de donaciones de apoyo se ejecutará en los mismos 25 distritos incluidos en el componente 1.

(b) Donaciones de innovación – son donaciones puestas a disposición de las comunidades para abordar problemas relacionados con la infancia. Con la donación de innovación se ayudará a implementar intervenciones fuera del menú de intervenciones descritas para las donaciones de apoyo a comunidades (a) más arriba. Tal como lo indica su nombre, el fondo de “innovación” se utilizará para apoyar a las comunidades en la ejecución de “ideas innovadoras” en diferentes niveles para mejorar la vida de los niños dentro de su comunidad. Las comunidades acceden a estas donaciones de la misma forma que para las donaciones de apoyo a comunidades: esto es, las comunidades elaboran propuestas siguiendo un ejercicio de planificación participativo, luego son seleccionadas por un comité a nivel de subcondado y finalmente remitidos para ser financiadas por el proyecto.

Componente 3 del proyecto– Programa Nacional de Apoyo para el Desarrollo Infantil. Este componente comprende actividades del programa central e iniciativas de políticas diseñadas para apoyar los programas de los componentes 1 y 2 a nivel de cada distrito y proporcionar una garantía de calidad para las actividades de primera línea del proyecto a nivel comunitario. En este componente se incluye (a) supervisión y evaluación del programa, (b) apoyo para la prevención de deficiencias de micronutrientes, (c) creación de currículum para

el desarrollo de la primera infancia, (d) capacitación de entrenadores en desarrollo de la primera infancia y (e) información, educación y comunicaciones.

I. Disposiciones para la implementación

La implementación del proyecto es responsabilidad del Gobierno de Uganda, con la asistencia de organizaciones no gubernamentales dentro del marco de descentralización y delegación de poderes a los niveles inferiores, según lo estipulado en las políticas nacionales. La comunidad (LC-1) es la unidad de operación para la prestación de los servicios, si bien la estructura de coordinación también involucrará a los niveles de subdivisiones de condado (LC-2), subcondados (LC-3) y distrito (LC-5).

Además, con el proyecto se espera hacer uso de actividades de consulta y sensibilización de las partes interesadas, movilización comunitaria, planificación participativa de la comunidad y creación de capacidades para la sostenibilidad en todos los niveles, junto con el fortalecimiento de la complementariedad con programas y estructuras nacionales que ya existan. Las estructuras políticas e institucionales existentes se deben usar en forma multisectorial. También se debe garantizar la transparencia y la responsabilidad pública en todos los niveles.

I. Coordinación del proyecto

Nacional. Un Comité de Dirección del Proyecto (PSC) compuesto por representantes de ministerios sectoriales, donantes y ONG estará a cargo de la orientación general de la implementación del proyecto.

El comité, que será encabezado por el Secretario Permanente del Ministerio de Finanzas, Planificación y Desarrollo Económico (MFPED), proporcionará orientación respecto a temas de políticas y revisará y aprobará la calidad y eficiencia de la implementación. El comité de dirección del proyecto también hará sugerencias para mejorar el presupuesto anual y planes de trabajo de los distritos para el proyecto.

Una pequeña Oficina de Coordinación del Proyecto (PCO), compuesta por un coordinador, un coordinador subrogante, un contador calificado y personal de apoyo poco numeroso se constituirá en el sector de servicios sociales del Ministerio de Finanzas, Planificación y

Desarrollo Económico y estará a cargo de la coordinación diaria de las actividades del proyecto a nivel nacional.

Distrito. El Director Ejecutivo Administrativo (CAO) identificará a un comité multisectorial –el Comité de coordinación de distritos (DCC)– para que se haga cargo de la responsabilidad de coordinar el proyecto a nivel de los distritos. El CAO identificará a una persona de enlace de entre los funcionarios gubernamentales que coordinará las actividades relacionadas con las ONG y otras actividades del proyecto. La ONG rectora será incluida como miembro del DCC. En aquellos distritos donde no sea posible identificar o fortalecer a una ONG con la capacidad y habilidad adecuada para asumir como ONG rectora, la implementación se llevará a cabo a través de la administración del distrito.

Subcondado. Un comité sectorial existente similar a aquel del nivel de distrito será responsable de coordinar las actividades del proyecto a nivel de subcondado. Esta comité de coordinación de subcondado también facilitará las relaciones entre las estructuras existentes y aquellas del proyecto y junto con la ONG rectora para el distrito aprobará los planes de trabajo anual y los requisitos de financiamiento de las ONG/CBO (Organizaciones comunitarias) para el proyecto en el subcondado.

I. Evaluación de impacto del proyecto

El Gobierno de Uganda está interesado en evaluar el impacto de diferentes aspectos del proyecto, para así determinar su eficacia y orientar el diseño de futuros proyectos de Nutrición y Desarrollo de la Primera Infancia. Más aún, puesto que el Banco Mundial considera este proyecto una posible fuente de información para otros países respecto a los servicios en cuestión, ha incluido al proyecto en una evaluación de tres países de programas de Desarrollo de la Primera Infancia y proporcionará asistencia técnica a modo de donación a la Oficina de Coordinación del Proyecto para promover actividades específicas de investigación. En particular, se llevarán a cabo dos estudios para evaluar el impacto de intervenciones específicas del proyecto como parte del proyecto general:

- Estudio del día de la salud infantil en subdivisiones de condado, para evaluar la cobertura de los tratamientos vermífugos suministrados en los días de salud infantil a nivel de cada subdivisión y su impacto en el aumento de peso de niños menores de seis años, usando un diseño experimental aleatorio.
- Investigación de encuestas usando metodología de línea base y de re-encuesta para evaluar (a) el impacto de los tratamientos vermífugos y de las actividades generales del proyecto en el desarrollo cognitivo, salud y nutrición de los niños menores de seis años;

(b) el impacto del componente de educación de personas encargadas del cuidado de los niños y la campaña de comunicaciones en los medios de prensa en cuanto al conocimiento, actitud y prácticas de crianza de los niños de los principales encargados; y (c) el impacto de la capacitación administrativa a nivel de base, de las actividades de generación de ingresos y formación de grupos de ahorro y crédito y la provisión de donaciones comunitarias en el bienestar del hogar y la comunidad.

La consultora seleccionada proporcionará el apoyo técnico y logístico para llevar a cabo los anteriores estudios y será invitada a participar como ejecutora local de investigación en el diseño, recopilación y análisis de datos necesarios para completar los dos estudios de evaluación de impacto. Esta consultora será la principal contrapartida de la Oficina de Coordinación del Proyecto, los investigadores locales y los investigadores del Banco Mundial y la Universidad de Oxford que llevarán a cabo la evaluación del impacto.

I. Descripción general de los estudios

Estudio uno: Impacto de desparasitamiento en los días de la infancia en subdivisiones de condado. Hay una gran cantidad de estudios sobre el impacto de tratar niños en edad escolar con medicamentos vermífugos. Sin embargo, sólo hay un ensayo aleatorio a gran escala que demuestra un gran efecto en el aumento del peso en niños preescolares. Debido a esto ha surgido la pregunta de si es posible lograr o no dicho efecto en los niños africanos. Por lo tanto, en el proyecto de Nutrición y Desarrollo de la Primera Infancia se ha incluido un estudio aleatorio sobre el impacto de entregar el agente antiparasitario, anthelmintic albendazole, a niños pequeños en 25 subdivisiones de condado seleccionadas en el transcurso de los días de la infancia en subdivisiones de condado y para medir el efecto de tratamientos semestrales en el aumento de peso. Se recopilarán datos de estas subdivisiones, al igual que de 25 grupos de control que también organizarán días de salud infantil, pero donde no se administrará albendazole en forma rutinaria. Si los tratamientos vermífugos se realizan con éxito y se demuestra su efecto benéfico en los niños ugandeses, entonces se puede recomendar este tratamiento a todos los distritos.

Debido a que se trata de un ensayo controlado científicamente, la selección de las subdivisiones a las que se les solicitará administrar el albendazole la llevará a cabo la Oficina de Coordinación del Proyecto (PCO) de una lista de subdivisiones de condado donde se organizarán días de la infancia (esta lista la entregarán las ONG que trabajan en las subdivisiones). La PCO también seleccionará subdivisiones que servirán como grupos de control. Este diseño experimental es clave para tener éxito en la evaluación.

La consultora se asegurará que las ONG locales responsables de la organización de los días de salud infantil en las subdivisiones estén conscientes de los fundamentos del diseño experimental y de que cumplan con la estrategia. Cada niño entre 12 meses y seis años que asista

a la feria en las 25 subdivisiones designadas recibirá una única tableta de 400 miligramos de albendazole masticable específica. El albendazole se administrará cada 6 meses; en caso de que las ONG decidan organizar días de la infancia con una mayor frecuencia, el antiparasitario se seguirá administrando cada seis meses y no más a menudo.

Los niños en las subdivisiones donde se administrará el albendazole, al igual que los niños en las 25 subdivisiones de control designadas, serán pesados en cada día de la infancia y sus pesos se registrarán tanto en su propia tarjeta de salud como en el registro comunitario. Aquellos niños demasiado pequeños para pararse en la balanza por sí solos, se pesarán en los brazos de sus madres después de fijar la balanza en cero con la madre parada sola sobre la balanza. Estos pesos se registrarán a la décima más cercana (0,1) de un kilo. Los datos en los registros comunitarios son responsabilidad de las ONG locales, si bien la consultora trabajará con éstas para asegurar que el sistema de recopilación de datos sea compatible con la gama completa de objetivos del estudio.

La consultora transcribirá estos pesos a una proforma que será diseñada en conjunto con los asesores técnicos del Banco Mundial y la Universidad de Oxford. Esta transcripción de los datos se llevará a cabo cada seis meses después del día de la infancia en el área del proyecto. Además de la identificación del niño (una combinación exclusiva de la identificación de la subdivisión, la aldea y la identificación individual registrada tanto en la tarjeta de cada niño como en el registro de la comunidad), los datos de la proforma incluirán el género del menor; la fecha de nacimiento que aparece en su tarjeta clínica o, en caso de no tener ésta, la edad del menor que aparece en los registros de la subdivisión; la fecha de la feria infantil en la que se registró el peso; y si el niño tomó o no una dosis de albendazole. Estos datos se ingresarán en un archivo computarizado en Kampala. La identificación individual proporcionará la base para fusionar los datos de diferentes períodos, por ello es necesario registrar la identidad cada vez que los datos se transcriben y ésta deberá ser constante para un niño durante todo el proyecto.

También se registrarán las condiciones y circunstancias locales en cada día de la infancia que puedan hacer desistir a las madres de acudir al evento. Esto incluye datos sobre el estado de las cosechas y las condiciones climáticas, ambas condiciones que pueden impedir la asistencia de las madres. Se registrará cualquier método y oportunidad especial aprovechado para publicitar cada día de la infancia, debido a que las diferentes formas de publicidad pueden influir en la asistencia. En el registro también se debe incluir una estimación de la cantidad de niños que asistieron al evento provenientes de otras subdivisiones y aquellos que no contaban con números de identificación otorgados por los organizadores.

El experimento tendrá una duración de dos años. Por lo tanto, la consultora registrará los datos cinco veces para cada subdivisión. Es decir, la consultora recopilará datos al inicio del proyecto y 6, 12, 18 y 24 meses después.

Se enviará una copia completa de los datos a la PCO cada 6 meses. Estas copias se considerarán los servicios prestados del primer estudio del proyecto. Se emprenderán análisis preliminares en la Universidad de Oxford también cada seis meses. Sin embargo, la consultora debe nombrar a un representante que participará en el análisis principal que se llevará a cabo una vez transcurrido los dos años. Se pagarán los gastos de viaje y estadía para que este representante trabaje en el análisis en Oxford. Los fondos para este viaje se consignan en el presupuesto como una partida separada y por lo tanto, no es necesario incluirlos en el contrato de la solicitud de propuesta.

Estudio dos: Impacto global de las intervenciones de nutrición y desarrollo de la primera infancia. En encuestas domiciliarias y comunitarias se recopilará la información básica y de seguimiento necesaria para evaluar el impacto de las diferentes actividades del proyecto. Las encuestas tendrán varios módulos, los que medirán:

- El desarrollo cognitivo y crecimiento de niños menores de seis años como resultado del tratamiento vermífugo y de las actividades generales del proyecto. En el Estudio dos se evaluarán los resultados de crecimiento longitudinal y desarrollo sicosocial y cognitivo de un grupo de niños de comunidades que participan en el proyecto (con y sin tratamiento vermífugo), en comparación con un grupo de niños de comunidades que no participan en el proyecto. Se hará un seguimiento a ambos grupos durante dos años o más. Con el estudio, por lo tanto, se complementará el estudio de desparasitamiento a nivel de cada subdivisión, permitiendo comprender mejor la decisión de llevar a los niños a los días de la infancia y para medir si la participación conduce o no en el tiempo a un aumento en las medidas de desarrollo cognitivo. Más aún, al incluir a las comunidades que no recibieron estos servicios de nutrición y desarrollo de la primera infancia, en el estudio se evaluará si el paquete de servicios produce mejorías en el estado nutricional y el desarrollo cognitivo.
- Cambios en los conocimientos, actitud y prácticas de crianza de las personas a cargo del cuidado de los niños, como resultado de la educación de los progenitores y la campaña en los medios de prensa.
- Mejoramiento en la salud y nutrición de niños menores de seis años como resultado de las actividades de supervisión del crecimiento, educación en salud preventiva y nutrición, tratamientos vermífugos y actividades generales del proyecto.
- Bienestar del hogar como resultado de donaciones comunitarias, capacitación en administración a nivel de base, actividades generadoras de ingresos y formación de grupos de ahorro y crédito.
- Características y cambios de la comunidad como resultado de las intervenciones del proyecto (u otras) que podrían influir en el bienestar de los niños durante el desarrollo del proyecto.

Selección de la muestra. La base para este estudio será una encuesta básica recopilada en el momento en que se presten los servicios por primera vez a las comunidades y una encuesta de seguimiento llevada a cabo en los mismos hogares dos años después de la encuesta inicial. Un tercio de la muestra provendrá de las mismas 25 subdivisiones del grupo con tratamiento (vermífugo) y otro tercio de los grupos de control del Estudio uno. Además, un tercio de la muestra provendrá de aldeas en 25 subdivisiones de los mismos distritos que los grupos de tratamiento, pero que no deberían recibir servicios del proyecto de nutrición y desarrollo de la primera infancia. Se seleccionarán 30 hogares de cada condado. Esto implica 750 hogares por estrato (2.250 en total) en la encuesta inicial. Dada el desgaste supuesto de la muestra, en la segunda encuesta se espera una disminución de entre un 5-10 por ciento de hogares.

Para recopilar la muestra en las subdivisiones de tratamiento y control, se hará una lista de todos los hogares en cada subdivisión (hay aproximadamente 700 hogares en una subdivisión, como promedio), posiblemente por un residente de la comunidad. Esta lista incluirá el nombre del jefe o jefa de hogar, una indicación de su ubicación y la cantidad de niños menores de seis años que contiene. Esta lista tiene dos objetivos. En primer lugar, se seleccionará una muestra de 30 hogares con por lo menos un niño menor de seis años por subdivisión, en un sorteo aleatorio. En segundo lugar, la cantidad total de niños menores de seis servirá como estimación de la posible cobertura de niños en los días de la infancia y, por lo tanto, ayudará a determinar la tasa de asistencia.

Debido a que el proyecto de nutrición y desarrollo de la primera infancia tendrá menos contacto con las comunidades en las que actualmente el proyecto no tiene actividades, la selección de hogares que no reciben estos servicios debería basarse en un muestreo por grupo para reducir los costos del listado de muestras. Específicamente, se debería seleccionar un subcondado que no esté en el proyecto por cada subcondado que sí esté en el grupo de tratamiento, preferiblemente uno adyacente. Se debe hacer una lista de todas las subdivisiones de este subcondado y seleccionar al azar 25 subdivisiones de la lista total. Luego se escogerán dos aldeas de cada subdivisión seleccionada, nuevamente con la ayuda de una lista de todas las aldeas de la subdivisión. Con este paso se reduce la cantidad de aldeas donde será necesario llevar a cabo una encuesta. El censo, similar al que se usa en las subdivisiones de tratamiento y control, constituirá la lista de hogares utilizados para extraer la muestra de 30 hogares por subdivisión. Esto formará el tercer estrato de la encuesta.

La encuesta básica inicial se debería realizar a mediados de 1999. La determinación de esta fecha se basa en la necesidad de conocer los subcondados y subdivisiones donde las ONG estarán funcionando, a fin de utilizar el diseño de muestra sugerido. La fecha también se basa en el supuesto de que la selección y capacitación de las ONG rectoras no estará lista antes de fines de 1998.

Sin embargo, la elaboración y pruebas preliminares al cuestionario deben llevarse a cabo mucho antes de esta fecha (a comienzos de 1999), para estar listos para implementar la encuesta apenas las ONG hayan identificado a las subdivisiones en las que trabajarán. Puesto que los

datos básicos necesitan ser llevados al terreno antes de realizar el primer desparasitamiento, el momento ideal para llevar a cabo la encuesta básica es simultáneamente con la primera organización comunitaria que realizará un día de la infancia. Dado que la muestra de 30 familias por subdivisión es pequeña en relación a la población total, es improbable que la recopilación de datos de la encuesta perturbe otras actividades o signifique una molestia para las comunidades. La recopilación de datos en los grupos de control (aquellos con actividad de ONG, pero sin desparasitamiento, y aquellos con ninguna de las dos) debe ser simultánea a la recopilación de datos en el grupo de tratamiento.

Instrumento de la encuesta. Los cuestionarios básicos que se utilizarán para el proyecto de encuesta son cuestionarios de hogares (que recopilan datos a nivel de cada hogar e individuos) y cuestionarios de comunidades.

I. Encuesta domiciliaria

Los datos de los hogares se recopilarán usando un cronograma precodificado. Se redactará basándose en el modelo de las Encuestas de Niveles de Vida utilizadas en más de 30 países. Un primer borrador será proporcionado por investigadores del Banco Mundial. Sin embargo, el instrumento se condensará para ajustarse a las necesidades específicas del proyecto y se adaptará a las condiciones locales usando grupos representativos; adicionalmente, la consultora llevará a cabo un procedimiento de prueba preliminar. El cuestionario de hogares contendrá módulos para recopilar la siguiente información:

1. Características socio-demográficas: una lista de los individuos que han residido en el hogar durante los últimos 12 meses, su edad y género, al igual que su nivel de escolaridad y tipo de empleo (si viene al caso). El formato de codificación indicará a los progenitores de todos los niños, si están presentes; si no lo están, indicará si aún están vivos. Se recopilará una lista detallada de activos para servir como indicador de la situación socioeconómica.
1. Conocimientos, actitud y prácticas: En el cuestionario también se recopilará información respecto a los conocimientos, actitudes y prácticas de crianza de los encargados principales del cuidado de los niños.
1. Datos antropométricos: Se registrarán los pesos a la décima más cercana (0,1) al kilo para todos los menores de seis años, usando balanzas digitales que serán provistas. Además, se registrará la altura de todos los menores entre dos y seis años. La prueba preliminar se utilizará para determinar si también es factible recopilar los pesos de las madres de estos niños (si viven en el hogar).
1. Evaluación cognitiva: La consultora trabajará con otras consultoras de investigación locales e internacionales de la Oficina de Coordinación del Proyecto para integrar pruebas de desarrollo cognitivo infantil en la recopilación global de los datos en terreno. En la

encuesta básica se tomará una prueba de desarrollo cognitivo reconocida internacionalmente a todos los menores entre 4,0 y 5,99 años. Esta prueba también se tomará al mismo grupo etéreo en la segunda ronda de la encuesta, permitiendo hacer una comparación de grupos. Además, se tomará esta misma prueba a un subconjunto de menores entre 6 y 7,99 años al momento de la segunda ronda (esta estrategia se resume en el Cuadro 2.1).

Además, se pueden diseñar evaluaciones de conocimientos en base a contenidos específicos del programa y una evaluación dinámica para la segunda ronda de la encuesta. La inclusión de estas medidas se evaluará durante el transcurso del proyecto. Por último, se creará una medición del rendimiento escolar para evaluar el conocimiento adquirido en el primer año de escuela y se administrará a un subconjunto de niños mayores en la segunda encuesta. Se pueden adaptar pruebas ya disponibles.

1. Salud infantil: Los datos sobre morbilidad (inclusive número y tipo de síntoma, niveles de gravedad, prolongación en el tiempo), patrones de acceso y utilización de los servicios de salud, servicios sanitarios y así sucesivamente.
1. Economía del hogar: se hará una amplia exploración del mejor enfoque para recopilar esta información durante la fase de la prueba preliminar y se evaluará en conjunto con los asesores de la Oficina de Coordinación del Proyecto antes de terminar el cuestionario. Entre las variables se puede incluir gastos en alimentación, actividades de agropastoreo, consumo de producción doméstica, gastos no en alimentos, características de la vivienda, inventario de bienes durables, situación laboral, actividades económicas, ingresos, tierra, cultivos y animales, ingresos de actividades del proyecto, empresas familiares, propiedad de activos, información sobre créditos y ahorro respecto a la cantidad de dinero y bienes prestados y que se adeudan, si se ha pedido dinero o bienes prestados en los últimos 12 meses, ahorros y endeudamiento neto el día de la entrevista, información sobre préstamos, inclusive el plan de pago, motivo para endeudarse y cantidad de préstamos de la misma fuente, y ubicación de los ahorros, si los hay, incluyendo banco, banco de ahorro para la vivienda, banco de ahorro rural, entre otros. Esta información formará parte sólo de los datos básicos y de las encuestas finales.

I. Encuesta comunitaria:

Se utilizarán cuestionarios comunitarios para recopilar información sobre las condiciones locales comunes a todos los hogares del área. Se hará una amplia exploración del mejor enfoque para recopilar esta información durante la fase de prueba preliminar y éste se evaluará en conjunto con los asesores de Oficina de Coordinación del Proyecto antes de terminar el cuestionario. Las variables podrían incluir lo siguiente:

1. Información demográfica: cantidad de hogares, total de habitantes, habitantes menores de seis años, grupos étnicos y religiones;
1. Información económica, incluyendo las principales actividades económicas y patrones de migración por motivos de trabajo;

1. Infraestructura: acceso a caminos, electricidad, agua por cañería, mercados, bancos y transporte público. Condición de la infraestructura local, como caminos, fuentes de combustible y agua, disponibilidad de electricidad y medios de comunicación;
1. Condiciones y prácticas agrícolas locales: tipo de cultivos que se producen dentro de la comunidad, con qué frecuencia y el ciclo de plantación y cosecha, cómo se vende normalmente la cosecha y datos cualitativos sobre precipitaciones, condiciones climáticas y el ciclo de estaciones;
1. Educación: cantidad y tipos de establecimientos preescolares, disposiciones formales e informales para la nutrición y el desarrollo en la primera infancia, distancia hasta las escuelas, números de cursos, tasas de matrícula (bruta y por género), asistencia, nivel de promoción por curso, servicios de salud y nutrición que presta la escuela (por ejemplo, programas de salud escolar, almuerzo);
1. Salud: tipo de establecimiento de salud y distancia y tiempo de viaje hasta los diferentes tipos de establecimientos de salud más cercanos (hospital, farmacia, posta de salud, etc.). Distancia y tiempo de viaje para acceder a los diferentes tipos de trabajadores de la salud más cercanos (médico, enfermera, farmacéutico, partera, trabajador de salud comunitaria, etc.); y
1. Otros: cantidad y tipo de ONG/organizaciones comunitarias locales activas, otros proyectos o intervenciones relativos a la infancia (por ejemplo campañas de vacunación públicas) y otros proyectos de desarrollo comunitario.

Sugerencias de personal para la encuesta. El personal de la encuesta debería estar conformado de la siguiente manera:

- Personal básico de la encuesta: compuesto por el administrador de la encuesta, el jefe en terreno, el administrador de datos y el personal que ingresará los datos, quienes serán responsables de la supervisión, coordinación y monitoreo general en terreno de las actividades de recopilación, ingreso y administración de los datos.
- Personal de la encuesta en terreno: las operaciones en terreno serán ejecutadas por equipos compuestos de un supervisor, dos (o tres) entrevistadores responsables del cuestionario principal y las mediciones antropométricas y un conductor. Un número similar de especialistas, que participarán en la administración de las pruebas de desarrollo cognitivo a los niños, será seleccionado y capacitado en conjunto con los expertos locales e internacionales.
- Coordinador para el ensayo aleatorio: el coordinador prestará su asistencia en el desarrollo de los instrumentos de recopilación de datos y en la capacitación de las ONG locales responsables de la organización de los días de la infancia en las subdivisiones, en relación con el diseño experimental, la recopilación y la transcripción de los datos.
Esta persona supervisará el ingreso de los datos y la administración del conjunto de datos del estudio y participará en el análisis principal que se realizará al final de estudio.

Organización del trabajo en terreno: La consultora participará en la formulación de los instrumentos de trabajo en terreno antes de la prueba preliminar de la encuesta y tendrá la principal responsabilidad por dicha prueba preliminar. Después de esta prueba, los cuestionarios serán rediseñados (en conjunto con investigadores del Banco Mundial) y traducidos a los idiomas locales.

La consultora trabajará con otras consultoras de investigación locales e internacionales seleccionadas por la Oficina de Coordinación del Proyecto para integrar las pruebas de desarrollo cognitivo infantil en la recopilación general de datos en terreno. El investigador local en nutrición y desarrollo de la primera infancia, con la asistencia de consultores internacionales, seleccionará y adaptará las principales pruebas cognitivas que se utilizarán para capacitar a las personas que tomarán dichas pruebas.

Se sugiere la siguiente organización del trabajo en terreno. Esta propuesta se basa en experiencia internacional y está diseñada para asegurar un control de la calidad. Es posible acordar ciertas variaciones al enfoque en consulta con los investigadores del Banco Mundial, basándose en la experiencia de la consultora y otros asesores de la Oficina de Coordinación del Proyecto y la información obtenida durante la prueba preliminar.

La organización del trabajo en terreno constará de pequeños equipos compuestos por un supervisor, dos (o tres) entrevistadores responsables del cuestionario principal y las mediciones antropométricas y una cantidad similar de especialistas para tomar las pruebas de desarrollo cognitivo a los niños. Este personal será capacitado en Kampala por el investigador local en desarrollo de la primera infancia, en coordinación con asesores internacionales sobre pruebas psicológicas. La capacitación incluirá un análisis de los objetivos que se persiguen con la investigación, una revisión de cada paso de la entrevista, capacitación práctica en la oficina, una marcha blanca en terreno y una recapitulación de la experiencia posterior a la marcha blanca.

Una vez que los equipos hayan sido capacitados, deberían retenerse durante toda la ronda de la encuesta, si es posible. Sin embargo, dado que durante el trabajo en terreno se podría descubrir que algunas personas son inadecuadas, se recomienda capacitar a algunas personas adicionales. No es aconsejable contratar personal para que trabaje sólo unos pocos días en una subdivisión y luego nuevo personal en otra subdivisión, porque con esto no se permite que el personal gane experiencia. Todo el personal debe recibir nueva capacitación al comienzo de la segunda encuesta.

Durante la administración de la prueba cognitiva, en la medida de lo posible los niños deben estar solos con el entrevistador. Bajo ninguna circunstancia se debe permitir que otra persona (adulto o niño) responda las preguntas que se le hacen al menor sometido a la prueba. Sin

embargo, durante la segunda encuesta, en la prueba del subconjunto de ocho años se puede utilizar un formato en grupo si resulta conveniente.

El supervisor será responsable de asegurar que los entrevistadores realicen la encuesta en los hogares escogidos para la muestra sin reemplazarlos y que se tomen las pruebas de desarrollo cognitivo a todos los niños de los grupos etáreos adecuados. Además, el supervisor revisará cada cuestionario una vez llenado (antes de que el equipo se traslade a otro condado), para asegurar que no queden espacios vacíos y verificar información que pareciera ser incoherente.

La consultora ingresará todos los datos de la encuesta tan pronto sea posible después de haber sido recopilados. Las copias de los datos de hogares y niños específicos y las escalas de clasificación, junto con los documentos necesarios para acceder a los datos, serán entregados a la Oficina de Coordinación del Proyecto en un formato computarizado al final de la encuesta básica. La consultora debe guardar los cuestionarios originales, porque por lo general será necesario acceder a los datos originales durante el transcurso del análisis.

Los datos de los niños deben contener códigos de identificación precisos que se puedan hacer coincidir con los códigos de la encuesta domiciliaria. Si bien los códigos especiales individuales y de hogares provistos por la Oficina de Coordinación del Proyecto no necesitan contener el nombre del hogar ni su ubicación exacta, la consultora debe guardar esta información de manera tal que sea posible volver a visitar el hogar en una fecha posterior. Debido a que en una etapa del análisis se deben vincular a los individuos de la segunda encuesta con sus resultados en las pruebas de la encuesta básica, todos los códigos de los individuos y hogares deben mantenerse constantes durante las tres encuestas.

I. Tareas específicas para los especialistas en encuestas

La consultora participará en las siguientes actividades en colaboración con la Oficina de Coordinación del Proyecto, investigadores locales e investigadores del Banco Mundial y de la Universidad de Oxford y ONG ejecutoras:

- Revisión de los programas de trabajo
- Desarrollo y adaptación de los instrumentos de recopilación de datos y la documentación de respaldo, inclusive materiales para hacer listas, cuestionarios, pautas de codificación, manuales para el entrevistador y el supervisor, manual de operaciones, manual de entrada de datos y procedimientos en terreno.

- Revisión de diferentes borradores de documentos, disposición, traducción, re-traducción y pruebas en terreno. Proporcionar copias impresas y electrónicas de todos los documentos a la Oficina de Coordinación del Proyecto. Enviar cuestionarios a los investigadores del Banco Mundial para que los revisen antes de la prueba piloto.
- Lista de viviendas y actualización cartográfica. Entre las responsabilidades por la confección de listados de hogares y viviendas en cada subdivisión seleccionada se incluye el obtener mapas básicos, elaborar materiales para hacer listas, contactar a las autoridades locales para informarles acerca de la operación de listado, identificar fronteras, dibujar mapas, enumerar hogares en forma sistemática, obtener información preliminar sobre los hogares, incluir el nombre del jefe o jefa de hogar, incluir una indicación de la ubicación del hogar y la cantidad de miembros del hogar menores de seis años; documentar los procedimientos al momento de diseñar la muestra, al final del trabajo en terreno y al completar el archivo de datos.
- Preparar el marco del muestreo (con el especialista en muestreo), capacitar al personal para implementar la muestra diseñada, supervisar la fase de ejecución para asegurar la calidad de la muestra seleccionada y proveer un informe detallado donde se describan todos los pasos involucrados en el diseño e implementación de la muestra.
- Previa consulta con el Banco Mundial, participar en la determinación de una estrategia adecuada para identificar grupos de comparación (esto es, subdivisiones que no forman parte del proyecto).
- Selección y capacitación del personal en terreno. Esta actividad abarca todas las actividades necesarias para crear materiales de trabajo y manuales para las personas involucradas en las tareas en terreno. Se requerirá capacitación para los entrevistadores, supervisores de entrevistadores, supervisores de equipos, personal de entrada de datos y personal para los trabajos antropométricos.
- Operación en terreno, inclusive disposiciones logísticas para la recopilación de datos y para obtener el consentimiento de los hogares e individuos; mantener un registro del hogar para el estudio.
- Producción de informes sobre la marcha de las actividades: La consultora preparará informes de la marcha de los trabajos en terreno (cada seis meses), con copia para la Oficina de Coordinación del Proyecto y el Banco Mundial. La consultora también deberá elaborar una descripción básica de la encuesta. Esta descripción debe incluir el contenido de la encuesta, el plan de muestreo y su implementación, al igual que las técnicas en terreno que se utilizarán. Se debe incluir un cuestionario completo y documentación básica a modo de anexos.
- Creación de un programa de entrada de datos usando software que pueda comprobar los rangos y coherencia de los datos y generar informes que indiquen los datos omitidos, datos fuera de los rangos aceptados y respuestas incoherentes.
- Depuración e introducción de datos, administración de bases de datos y planes de tabulación, inclusive la creación de un programa de entrada de datos, manual de entrada de datos, capacitación para el operador de entrada de datos, comprobaciones de calidad de los datos

y directrices para usar los datos. Además, codificación de preguntas abiertas, verificación de datos, revisión de datos antropométricos con respecto a tablas de referencia estándar.

- Hacer cumplir el acuerdo sobre políticas para la utilización de los datos: La consultora y los investigadores involucrados en el proceso de recopilación y análisis de datos firmarán un memorándum de entendimiento con la Oficina de Coordinación del Proyecto estableciendo explícitamente la política respecto a temas como acceso a los datos, usuarios destinatarios, procedimientos para obtener copias de los conjuntos de datos y documentación y reglas sobre publicación y autoría.
- Realizar análisis de datos: La consultora llevará a cabo análisis de datos exploratorios (por ejemplo, frecuencias, tabulaciones por porcentajes y tabulaciones cruzadas) de las variables clave de la encuesta y sus correlacionadas. La consultora realizará un modelaje estadístico moderno de los impactos después de la segunda y tercera ronda, para determinar el avance general de los indicadores sociales (por ejemplo, nutrición, salud, ingresos y desarrollo comunitario) y los factores responsables de los cambios o la falta de ellos.
- Producir informes de análisis: La consultora informará los resultados después de la segunda y tercera ronda de las encuestas sobre la base de los análisis de los indicadores sociales y las covariadas. La consultora se coordinará con la Oficina de Coordinación del Proyecto y con el Banco Mundial respecto al Estudio del día de la salud infantil en las subdivisiones y a la recopilación del impacto en el desarrollo cognitivo, pero no estará a cargo de los informes finales sobre los resultados de estos estudios.

Entre las tareas específicas para la encuesta comunitaria se incluyen las siguientes:

- Trabajar con los asesores de la Oficina de Coordinación del Proyecto en la creación del cuestionario comunitario y explorar a profundidad, en la fase de la prueba preliminar, el mejor enfoque para recopilar esta información;
- Trabajar estrechamente con los organismos ejecutores (ONG rectoras y locales) en la recopilación de los datos de la comunidad;
- Contactarse con autoridades locales y dirigentes de las comunidades para explicarles el enfoque frente a la evaluación de impacto del proyecto y obtener el consentimiento de la comunidad para la investigación de la encuesta y el estudio de los días de la salud infantil;
- Entrevistar a informantes clave y obtener mapas, listas y otros registros de la comunidad;
- Obtener listas de los establecimientos de salud y educación (establecimientos preescolares y primarios), inclusive ubicación geográfica, área de cobertura y tipo de establecimiento (por ejemplo, privado o público);
- Obtener información demográfica de la comunidad, entre otros cantidad de hogares y habitantes por edad y género; y
- Obtener otros datos necesarios para los cuestionarios de la comunidad.

Entre las tareas específicas para el estudio de los días de la infancia se incluyen las siguientes:

- Participación en la creación del protocolo del estudio;
- Creación de instrumentos de recopilación de datos;
- Capacitación de ONG locales a cargo de organizar los días de la infancia en las subdivisiones para el diseño experimental;
- Supervisión de los datos recopilados durante el día de la infancia;
- Transcripción de los datos;
- Ingreso y administración de los datos; y
- Participación en el análisis principal que se llevará a cabo al final del estudio.

Cuadro 2.1 del Anexo: Propuesta de tamaños de las muestras para la evaluación de impacto del Proyecto de Nutrición y Desarrollo de la Primera Infancia, Uganda

Categoría	Desparasitamiento y educación de progenitores		Sin desparasitamiento y con educación de progenitores		Sin desparasitamiento ni educación de progenitores		Total
	Básico	Segunda ronda ^a	Básico	Segunda ronda	Básico	Segunda ronda	
Nº de subdivisiones	25		25		25		
Momento	Básico	Segunda ronda ^a	Básico	Segunda ronda	Básico	Segunda ronda	
Nº de hogares	750	700	750	700	750	700	2250
Nº de niños pesados en Día de Infancia ^b	5000	5000	5000	5000			20000
Nº de niños entre 0-5,99 con antropometría en hogares (media 2 por familia) ^c	1500	1395	1500	1395	1500	1395	11580
Nº de niños sometidos a pruebas cognitivas: prueba a todos los niños entre 4,0-5,99 en los hogares	500 ^d	465 ^e	500	465	500	465	2895
Nº de niños sometidos a prueba cognitiva y antropometría entre 6,0-7,99		Subconj. ^f	-	Subconj.	-	Subconj.	Subconj.
Tasas de matrícula escolar	25 comun.	25 comun.	25 comun.	25 comun.	25 comun.	25 comun.	

- a. Suponiendo una pequeña pérdida debido a desgaste de 8 por ciento en dos años
- b. Suponiendo que aproximadamente 200 niños asistirán a cada día de la infancia.
- c. Se suponen dos niños por familia, pero se reclutarán a las familias que tengan *cualquier* número de niños menores de seis. Aquí familia se refiere a una pareja formada por una madre (o su reemplazante) y un menor.
- d. Este es el máximo; el número real puede variar de acuerdo al número de niños entre cuatro y cinco años que se encuentren.
- e. Suponiendo la misma pérdida de 8 por ciento en dos años; sólo se tomarán pruebas a niños cuyos progenitores fueron entrevistados.

f. el número será un subconjunto de niños en el rango de edad cuyos progenitores fueron entrevistados. Estarán vinculados con la puntuación anterior. Aun si el número de niños a los que se les tomen pruebas aumentara en la segunda ronda, el tiempo para las entrevistas podría disminuir, puesto que hay bastante información que puede no ser necesario evaluar nuevamente. También es posible que el tamaño de este grupo sea reducido.

Fuente: Documento del proyecto del Banco Mundial.

Estudio de validez. Además de lo anterior, se agregará un pequeño estudio longitudinal para examinar la validez predictiva de la medida preescolar para el rendimiento escolar al final del primer año de escuela. En la encuesta básica se administrará una prueba a dos niños por comunidad entre 6,0 y 6,9 (que aún no asistan a la escuela), para $N = 150$. Estos niños se ubicarán cuando se realice la prueba posterior y se les administrará una prueba de rendimiento escolar dos años después, cuando tengan entre 8,0 y 8,99 años.

Programa de trabajo: El cronograma preliminar para el programa de trabajo es el siguiente:

- Mes 1. Inicio del proceso de construcción de indicadores de desarrollo cognitivo en conjunto con los consultores internacionales y en conformidad con los términos de referencia. Este proceso puede demorar hasta seis meses.
- Mes 2. Prueba preliminar inicial y revisión del cuestionario.
- Mes 5. Comienzo del listado de hogares para la selección de la muestra. Este paso depende de la selección de las ONG rectora y locales. No se puede llevar a cabo hasta que la Oficina de Coordinación del Proyecto y las ONG escojan las subdivisiones donde se organizarán los días de la infancia y luego seleccionen los lugares para el programa inicial de desparasitamiento. Al mismo tiempo, el cuestionario se debe traducir y probar nuevamente en terreno.
- Mes 7. Comienzo de la recopilación de datos en ferias infantiles para el estudio de desparasitamiento. Los datos se recopilarán en estas ferias cada seis meses. Al igual que lo anterior, el momento para realizar este paso depende de la selección de las ONG rectora y locales.
- Mes 8. Capacitación del personal en terreno para la encuesta domiciliaria e inicio de la encuesta. La encuesta tomará aproximadamente entre tres y cuatro meses, dependiendo de la cantidad de equipos que se usen. La entrada de datos se debe llevar a cabo simultáneamente con la recopilación de los mismos.
- Mes 14. Análisis inicial de datos básicos. Este será un proceso permanente.
- Mes 20. El personal de la consultora visitará la Universidad de Oxford para participar en el análisis de los datos iniciales.
- Meses 20-36. Recopilación de datos para la segunda ronda, para el estudio de desparasitamiento.

Las encuestas domiciliarias a mediados del período y final se llevarán a cabo dos y cuatro años después del inicio.

Apoyo a la consultora

No se entregará apoyo específico a la consultora para llevar a cabo las tareas encomendadas. Se les recomienda a las consultoras incluir en sus propuestas todos los requisitos para llevar a cabo las asignaciones en forma eficiente.

Ejemplo II: Evaluación del impacto de los caminos rurales: Datos básicos de 1997 para Vietnam *

Términos de referencia: Encuesta básica para el estudio del impacto de los caminos rurales

I. Antecedentes

En el estudio se pretende evaluar el impacto en los niveles de vida del proyecto de transporte rural de Vietnam financiado por el Banco Mundial, el que fue ejecutado en 15 provincias pobres durante tres a cinco años a partir de 1997. El punto central general del estudio estará puesto en el cambio de las determinantes de los niveles de vida en el tiempo en aquellas comunas que tienen intervenciones de proyectos viales, en comparación con aquellas que no las tienen. Esto exige la recopilación de datos básicos previos al proyecto tanto para las áreas del proyecto (“tratamiento”), como para las áreas de control sin tratamiento, además de un número de otras rondas de recopilación de datos posteriores a la intervención en intervalos de dos años. Se creará una base de datos detallada a nivel de cada comuna, en parte basándose en registros recopilados anualmente en cada comuna. A esto último se le agregará la recopilación de datos comunales retrospectivos y la recopilación de diferentes datos complementarios clave. Una breve encuesta distrital ayudará a contextualizar los datos comunales. Por último, se hará un muestreo aleatorio de entre 10 y 15 hogares de las listas de hogares comunales y se aplicará un breve cuestionario de hogares. El estudio se llevará a cabo en 6 provincias de las 15 que se beneficiarán del proyecto. Las seis provincias serán representativas de las 6 regiones geográficas de Vietnam. Se extraerá una muestra aleatoria de aproximadamente 200 comunas con y sin proyectos. Se conformarán seis equipos para entrevistar simultáneamente a cada provincia. La encuesta debe comenzar en abril y concluir aproximadamente en agosto. Los datos deben estar disponibles entre octubre y noviembre.

I. Diseño de la encuesta

Muestreo: El muestreo se llevará a cabo en tres niveles:

* Estos términos de referencia fueron elaborados por Dominique van de Walle.

1. Provincias: Las 15 provincias que conforman el proyecto están ubicadas en las 6 regiones geográficas de Vietnam. Los criterios para seleccionar las provincias incluidas en la encuesta serán los siguientes: (a) se seleccionará una provincia en cada región geográfica; b) cuando en cada región haya más de una provincia con posibilidades de ser incluida en el proyecto, se hará una selección aleatoria.

2. Comunas: El objetivo es entrevistar 200 o más comunas, las que se seleccionarán aleatoriamente. Aproximadamente la mitad o menos (no más) deben ser comunas con proyectos de enlaces viales, el resto serán controles. Se compilará una lista de las comunas sin proyectos en las seis provincias (o alternativamente una lista para cada provincia) y se extraerá una muestra aleatoria. De igual modo, se compilará una lista de todas las comunas que se benefician de proyectos viales en las seis provincias (o por provincia). Esto puede ser más de una comuna por enlace vial; todas se incluirán en el marco del muestreo. De estas comunas también se extraerá una muestra aleatoria. La muestra no necesariamente incluirá ambas comunas vinculadas por un proyecto vial. Si el acceso a ciertas comunas de la muestra es imposible, serán reemplazadas por otra comuna similar dentro del distrito

3. Hogares: En cada comuna de la muestra se aplicará una encuesta a 15 hogares. Éstos (más tal vez algunos hogares de reemplazo) se seleccionarán al azar de las listas de hogares de las comunas. Después de la selección, se le pedirá a las autoridades comunales que clasifiquen a los hogares como muy pobres, pobres, promedio, no pobres y ricos.

I. Procedimiento de la encuesta

Se contratará a seis expertos en encuestas para llevar a cabo las encuestas en las 6 provincias. Después de capacitarlos y de probar el cuestionario en terreno, comenzarán a aplicar simultáneamente la encuesta en cada provincia. En los distritos, los encuestadores necesitarán por lo menos a un funcionario local de la unidad de administración distrital del proyecto para que ayude en el contacto con las autoridades locales y, en algunos casos, a encontrar guías e intérpretes adecuados para ciertas áreas menores. Se contratarán asistentes de encuestas o asistentes de las unidades provinciales de administración del proyecto, según sea necesario.

Cada encuestador recopilará datos de 35 comunas como promedio, los distritos a los que pertenecen y aproximadamente 15 hogares por comuna. Se necesitarán tres a cuatro días en cada comuna. El trabajo en terreno demorará entre 100 a 140 días (cuatro a cinco meses). El tiempo total será de seis meses.

Durante el tiempo de ejecución de la encuesta, el supervisor llevará a cabo visitas a terreno en las seis provincias para supervisar la recopilación de datos y asegurar su buena calidad.

Los datos recopilados se depurarán e ingresarán usando un programa de ingreso de datos.

En el Cuadro 2.II.1 del Anexo se muestra un cronograma estimado para el estudio.

Cuadro 2.II.1 del Anexo Cronograma para el estudio de la evaluación del impacto, Vietnam

	<i>Ene</i>	<i>Feb</i>	<i>Mar</i>	<i>Abr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Ago</i>	<i>Sep</i>
Diseñar cuestionarios	****	****							
Probar encuesta en terreno		****							
Revisar cuestionarios			****						
Adaptar programa de entrada de datos, traducir e imprimir los cuestionarios			****						
Contratar y capacitar a encuestadores			****						
Encuesta en terreno				****	****	****	****	****	
Comprobación de datos								****	****
Entrada de datos									*****

I. Otros

Equipos: Los equipos adquiridos en el marco del proyecto serán de propiedad del proyecto mientras el estudio esté en ejecución (hasta futuras rondas), pero cuando el equipo de trabajo no los esté usando, éstos se guardarán en la Unidad de Administración del Proyecto y estarán disponibles para su uso.

Desembolsos presupuestarios. El presupuesto para el estudio (excluyendo pagos al investigador principal, quien recibirá pagos mensuales) se desembolsará en tres asignaciones. El primero, al firmar el contrato, consistirá en el 20% del total de fondos. La segunda asignación, que corresponderá al 50% del presupuesto total, se desembolsará una vez que se terminen los cuestionarios para las comunas, hogares y distritos y hayan sido aprobados por el jefe de proyecto del Banco Mundial. Esto debería suceder en algún momento a fines de marzo. La tercera y última asignación se desembolsará a fines de julio o entremedio de la recopilación de datos. En el Cuadro 2.II.2 del Anexo se muestran los detalles del presupuesto estimado.

Cuadro 2.II.2 del Anexo Estimación del presupuesto para el estudio

	<i>No.</i>	<i>Plazo</i>	<i>Monto (US\$)</i>	<i>Total (US\$)</i>
1. Investigador principal	1	9 meses	\$1.000	\$9.000
2. Expertos en encuestas	6	6 meses	\$400	\$14.400
3. Viático para 6 entrevistadores, 6 guías/interpretes locales	12	125 días	\$8	\$12.000
4. Automóvil y otro transporte para 6 equipos de encuesta	6	125 días	\$40	\$30.000
Arriendo de automóvil para investigador principal	1	30 días	\$50	\$1.500
5. Pasajes aéreos Hanoi-Hochiminh-Hanoi	6		\$200	\$1.200
Para encuestadores (provincias del sur)	3 personas			
Para el investigador principal	3 viajes			
6. Capacitación de encuestadores	12	1 semana	\$50	\$1.338
Pago		3 días/ 3	\$50	
Viaje a terreno		automóviles		
Viático		3 días	\$8	

7. Prueba en terreno del cuestionario (Comunas del sur y norte)	1	2 semanas		\$2.000
8. Depuración y entrada de datos	2	2 meses	\$200	\$800
9. Materiales de encuesta				\$2.000
10. Comunicaciones (fax, teléfono, e-mail, xerox)				\$2.000
11. Equipos				\$5.000
Computadora (PMU18)	1		\$1.700	
Impresora (PMU18)	1		\$ 500	
Fax (equipo del estudio)	1		\$1.800	
Computadora portátil (equipo del estudio)				
12. Traducción (cuestionario, manuales, documentación)	200 páginas		\$8/página	\$1.600
13. Impresión, xeroxado				\$800
14. Contingencias				\$1.362
Total				\$85.000

Términos de referencia: Supervisor de la encuesta o investigador principal

I. Descripción de la asignación

El supervisor de la encuesta o investigador principal en el país estará a cargo de llevar a cabo la encuesta básica del estudio dentro de Vietnam. Entre sus responsabilidades se incluye el determinar la disponibilidad de información a nivel comunal, ayudar a revisar y completar los cuestionarios a nivel de hogar, comuna y distrito, probar el cuestionario en terreno, incorporar revisiones al cuestionario, disponer la traducción del cuestionario, contratar y capacitar a los asistentes, planificar la logística del trabajo en terreno, preparar la ejecución de la encuesta y la documentación del cuestionario, supervisar la implementación de la encuesta y asegurar el control de la calidad y supervisar la base de datos del proyecto y disponer la depuración e ingreso de los datos. Esta persona también se desempeñará como enlace entre el la Unidad de Administración del Proyecto del Ministerio de Transportes PMU18 y la misión residente del Banco Mundial, el representante de la Agencia Canadiense de Desarrollo Internacional (CIDA) en Hanoi y el jefe de proyecto del Banco Mundial en Washington. Esta persona dependerá directamente del jefe de proyecto y asumirá sus funciones a partir de enero de 1997; el contrato se puede acordar por un período de nueve meses, por un monto de \$1.000 por mes.

I. Tareas específicas

Entre las tareas específicas se incluye lo siguiente:

1. Asumir la responsabilidad de contratar, redactar términos de referencia, capacitar y supervisar seis asistentes principales que trabajarán con asistentes locales (posiblemente de la oficina de transporte local) en terreno y estarán a cargo de recopilar datos a nivel de hogar, comuna y distrito;
1. Explorar la disponibilidad de datos a nivel de comuna y trabajar estrechamente con el jefe de proyecto del Banco Mundial para diseñar las versiones finales de los cuestionarios;

1. Ejecutar una prueba en terreno de los cuestionarios, tanto en comunas rurales del Sur como del Norte; informar sobre posibles problemas y revisiones necesarias; revisar los cuestionarios cuando sea necesario;
1. Disponer la traducción, impresión y reproducción (Xerox) de los cuestionarios (las versiones finales de los cuestionarios estarán disponibles tanto en inglés como en vietnamita);
1. Escoger las seis provincias que se incluirán en la encuesta, de modo que haya una provincia representando a cada región geográfica; cuando exista más de una provincia por región, la provincia para el muestreo se escogerá aleatoriamente; compilar una muestra aleatoria de aproximadamente 200 comunas rurales en las seis provincias, más o menos la mitad con proyectos y el resto sin proyectos;
1. Planificar toda la logística del trabajo en terreno, inclusive organizar el transporte, conductores, viáticos para viajes, programación de las encuestas en las comunas y notificar a las autoridades pertinentes de la llegada de los equipos y sus objetivos;
1. Participar en la implementación de la encuesta, alternando entre los equipos en una función supervisora; asegurar el control de calidad; identificar problemas que afecten la implementación de la encuesta; revisar la calidad y exhaustividad de los datos recopilados, sugerir maneras para solucionar problemas e implementarlas previa consulta con el jefe del estudio;
1. Asegurar que en las futuras rondas de la encuesta se pueda replicar la encuesta básica, lo que requiere (a) elaborar documentación detallada de todo el diseño y logística de la implementación de la encuesta (cómo se hizo el muestreo de las provincias, comunas y hogares; cómo se capacitaron y organizaron los equipos de encuesta; cómo se organizó en trabajo en terreno; qué procedimientos se siguieron cuando no se pudo acceder a un lugar de la muestra o no se encontró un hogar de la muestra; problemas, cuestiones que surgieron y soluciones aplicadas); y (b) redactar un manual detallado con definiciones de términos (por ejemplo desempleo, ingreso, ocupación principal, niño o adulto, distancia), unidades, montos en divisas, códigos utilizados en el cuestionario; cómo se deben aplicar los cuestionarios y a quiénes; cómo se recolectaron los precios, etc.; lo primero es para asegurar que las futuras rondas de la encuesta puedan reproducir la organización básica y los detalles logísticos y lo último se debe usar para capacitar a los entrevistadores y para su trabajo, al igual que para ayudar a futuros usuarios de los datos (habrá versiones en inglés y en vietnamita);
1. Conseguir los equipos necesario según el detalle del presupuesto del estudio;
1. Establecer buenas relaciones y asegurar una estrecha cooperación con PMU18. Mantenerlos al tanto del estudio y supervisar los acontecimientos del proyecto; supervisar la constitución de una base de datos con la información específica del proyecto (el jefe de proyecto del Banco Mundial identificará los datos que se incluirán);
1. Organizar y supervisar la depuración e ingreso de los datos usando el programa de entrada de datos provisto; y
1. Actuar como enlace y comunicarse a menudo con el jefe de proyectos.

Anexo 3

Muestra de presupuesto de la evaluación del impacto de un programa de alimentación escolar

Fase I: Julio de 1999-Diciembre de 2000^a

Propuesta de investigación sobre alimentación escolar– Evaluación básica y transversal
(Julio de 1999 - Diciembre de 2000)

Proyecto de presupuesto - 14/7/99 –US\$

	<i>Semanas personal /Actividad</i>		<i>Origen de fondos /Costos</i>			<i>Total</i>
	<i>AF00</i>	<i>AF01</i>	<i>BB</i>	<i>RPO</i>	<i>Otros</i>	
Personal del Banco Mundial						
Economista	4	2	17.640			
Especialista en evaluación	5	3	23.520			
Especialista en nutrición	5	3	23.520			
Colega revisor	0,2	0,2	1.948			
Colega revisor	0,2	0,2	1.948			
						68.577
Personal del FES						
Coordinador del estudio	4	4			12.000	
						12.000
Consultores internacionales						
Evaluación de la situación (incl. viaje)					7.000	
Elaboración prueba cognitiva (incl. viaje)				6.000		
Especialista en muestreo				2.000		
Estudio de eficacia-costos				25.000		
						40.000
Consultora regional^b						
Diseño, muestreo, administración				42.000		
Trabajo en terreno				25.000		
Procesamiento de datos				3.500		
Análisis				30.000		
						100.500
Viaje al país						
Viajes	4	2		12.000		
						12.000
Contingencias						
Comunicaciones				1.000		
Software				2.000		
Traducción				2.000		
						5.000
TOTALES			68.577	150.500	19.000	238.077

Total solicitado de RAD: \$150.500

Total solicitado del presupuesto del Banco: \$68.577

Total provisto por fuentes externas: \$19.000

a. Las estimaciones del presupuesto para la fase II de la evaluación no se incluyen en esta propuesta

b. En la página siguiente se desglosan estos costos

Presupuesto estimado – Recopilación y análisis de datos locales para la Fase I

Evaluación del impacto de la alimentación escolar -

Costos en US\$

	<i>Nº personas</i>	<i>Nº semanas de personal</i>	<i>Índice???</i> <i>semanal</i>	<i>Total, \$</i>
Profesionales				
Director	1	12	2.000	24.000
Especialista en educación	1	8	1.500	12.000
Especialista en nutrición	1	8	1.500	12.000
Estadístico/Muestreo	1	12	750	9.000
Ejecutivo en terreno	1	8	750	6.000
Programador	1	10	300	3.000
Supervisor del procesamiento de datos	1	8	300	2.400
Asistente – encuestas	1	10	100	1.000
Asistente – antropométrica	1	10	100	1.000
Asistente – pruebas cognitivas	1	10	100	1.000
Control calidad de datos	1	8	100	800
Sub Total - Personal profesional				72.200
Personal de trabajo en terreno				
Supervisor	4	6	200	4.800
Examinador prueba cognitiva	4	6	120	2.880
Experto mediciones antropométricas	4	6	120	2.880
Entrevistador	4	6	120	2.880
Conductor	4	5	100	2.000
Equipo de trabajo en terreno				
	Personas/unidades		Costo por semana o unidad	
Vehículos (4 vehículos durante 5 semanas)	4	5	350	7000
Gasolina (4 vehículos durante 5 semanas)	4		5	80
Balanzas; reglas (5 juegos)	5			20
Equipo prueba cognitiva (para 4 examinadores)	4			20
Equipo encuesta (para 4 entrevistadores)	4			20
Subtotal – Trabajo en terreno				24.300
Procesamiento de datos				
	Personas			
Codificación de datos	3		7	75
Entrada de datos	4		7	75
Subtotal - procesamiento de datos				3.675
Total				100.175

Anexo 4

Indicadores de impacto — Evaluación del Fondo de Inversión Social de Bolivia

Creado en noviembre de 1997

I. Educación formal -- Escuelas tipo "A" y "B" (multigrados y corrientes)

1. Indicadores del impacto final

Logros en pruebas de matemáticas y lenguaje^a

Índice de repitencia

Índice de deserción

Matrícula

Nivel de instrucción

Demanda por educación (% de estudiantes rechazados de la escuela)^a

2. Indicadores del impacto intermedio

Regularidad en la asistencia de los estudiantes

Regularidad en la asistencia de los maestros

Tiempo/horas asignadas por los estudiantes a estudiar

Método de enseñanza en aula^a

Rotación del personal docente^a

3. Indicadores de la intervención

Infraestructura

Relación estudiantes/salón de clases

Cantidad de salas de clase en "buen estado"

Cantidad de salas de clases que faltan

Disponibilidad de área multifuncional

Disponibilidad de servicios básicos

- Electricidad

- Fuente del principal suministro de agua

- Tipo de servicio sanitario; condición del servicio sanitario

Mobiliario

Relación estudiantes/pupitres

Relación pupitres de maestros/sala de clases

Relación sillas de maestros/sala de clases

Relación "pizarrón adecuado"/sala de clases

Relación estantes/salas de clases

Textos y materiales didáctico

Relación textos/estudiantes

Calidad de textos de matemáticas, lenguaje, ciencias sociales y naturales

Disponibilidad de libros para el maestro

Disponibilidad y condición de mapas y gráficos

Juegos didácticos por ciclo escolar (prebásico, básico e intermedio)
 Disponibilidad de un ábaco
 Indicadores de la reforma educacional^b

4. Factores que afectan los resultados no vinculados con el Proyecto FIS (Exógenos)

Nutrición
 Disponibilidad de programa de desayuno escolar
 Costo de la escuela
 Características de los maestros
 Antecedentes educacionales
 Años de servicio
 Capacitación recibida
 Métodos aplicados en la docencia (en un período de clases)
 Capacitación recibida, por asignatura y curso
 Prácticas de evaluación de los estudiantes (frecuencia de deberes y su corrección)
 Evaluación de los maestros por los estudiantes
 Fundamentos para la deserción escolar
 Estudiantes rechazados por la escuela
 Distancia entre el hogar y la escuela
 Relación estudiantes/maestro

5. Indicadores de identificación

Si la escuela fue priorizada por la Reforma educacional
 Costo programado por componente del proyecto
 Gastos efectivos por componente del proyecto

II. Salud

1. Indicadores del impacto final^c

Tasa de mortalidad infantil
 Tasa de mortalidad en la niñez
 Tasas de incidencia y prevalencia de las principales enfermedades
 Prevalencia de la desnutrición (general, leve, moderada y grave)

2. Indicadores del impacto intermedio

Uso de centros de salud del gobierno
 Prevalencia de la vacunación contra el tétano
 Lugar donde se recibió la vacuna
 Prevalencia de control prenatal
 Número de controles prenatales
 Calidad del control
 Prevalencia de partos asistidos en centros de salud
 Calidad de la atención
 Prevalencia de partos en el hogar asistidos por personal médico
 Talla al nacer
 Peso al nacer

- Evaluaciones antropométricas
 - Lugar donde se hace la evaluación
 - Edad en que se hace la primera evaluación
- Incidencia de enfermedades y prevalencia de inmunización según número de dosis recibidas
 - Polio
 - Difteria, Tétanos, Tos Convulsiva (DPT)
 - Sarampión
 - Vacuna contra la tuberculosis (BCG)
- Información sobre los lugares donde acudir por una vacunación
- Incidencia y tratamiento para la tos
- Incidencia y tratamiento para la diarrea
- Prevalencia de la información y uso de paquetes de rehidratación oral
- Conocimiento de las clínicas sobre la prevalencia de embarazos
- Asistencia de embarazos de alto riesgo
- Prevalencia de buenos hábitos de higiene y uso del agua
- Duración de la lactancia

3. **Indicadores de la intervención^d**

- Calidad de la infraestructura por tipo de centro de salud
- Disponibilidad de servicios básicos en el centro de salud (agua potable, sistema de alcantarillado y electricidad).
- Suficiencia de infraestructura en base a las normas establecidas por tipo de centro de salud
- Suficiencia del equipo en base a las normas establecidas por tipo de centro de salud
 - Número de camas en el centro de salud
- Disponibilidad de medicamentos esenciales por tipo de centro de salud
- Disponibilidad de instrumentos médicos esenciales por tipo de centro de salud
- Disponibilidad de mobiliario esencial por tipo de centro de salud

4. **Factores que afectan los resultados no vinculados con el Proyecto FIS (Exógenos)**

- Características del hogar
 - Calidad del hogar
 - Tipo de hogar
- Servicios básicos en el hogar
 - Electricidad
 - Fuente de agua
 - Tipo de servicios sanitarios
- Accesibilidad a los servicios básicos
 - Distancia entre el hogar y el centro de salud más cercano
 - Distancia entre el servicio sanitario y la fuente de agua
 - Distancia entre el hogar y la principal fuente de agua
 - Horas de disponibilidad de agua por día
 - Suficiencia de la cantidad de agua por día
 - Disponibilidad de agua a través del año

Costo de la consulta en el centro de salud
 Percepción del jefe/a de hogar respecto a la calidad de:
 - el "servicio" en el centro de salud al que asisten los miembros del hogar
 - la "infraestructura" del centro de salud al que asisten los miembros del hogar
 - la "disponibilidad de medicamentos" en el centro de salud al que asisten los miembros del hogar

Gastos familiares

Características personales de los miembros del hogar

Edad

Idioma

Nivel de educación

Ocupación

Características geográfica

Distrito de salud

Área de salud

Sector de salud

Provincia

Localidad

Recursos humanos en el centro de salud (doctores, odontólogos, nutricionistas, enfermeras, asistentes de enfermería, técnicos, personal administrativo)

Población en el área de influencia del centro de salud por grupos étnicos

Costo de la consulta en el centro de salud

Intervenciones de salud no financiadas por el FIS

5. Indicadores de identificación

Costo programado por componente de proyecto

Gastos reales por componente de proyecto

III. Agua

1. Indicadores del impacto final^e

Tasa de mortalidad infantil

Tasa de mortalidad de niños

Tasas de incidencia y prevalencia de diarrea en los hogares

Prevalencia de desnutrición (general, leve, moderada y grave)

2. Indicadores del impacto intermedio

Incidencia y tratamiento de la diarrea en centros de salud

Prevalencia del uso y conocimientos sobre el uso de paquetes de rehidratación oral

Prevalencia de buenos hábitos de higiene y uso del agua

3. Indicadores de la intervención (cantidad suministrada)

Prevalencia de capacitación en temas de salud

Accesibilidad a los servicios básicos

Fuente principal de agua

Existencia y tipo de servicio sanitario.
 Distancia entre el servicio sanitario y la fuente de agua
 Distancia entre el hogar y la fuente principal de agua
 Horas de disponibilidad de agua al día
 Suficiencia de la cantidad del agua durante el día
 Disponibilidad de agua durante el año
 Cantidad de agua que consume el hogar^a
 Calidad del agua^a

4. Factores que afectan los resultados no vinculados con el Proyecto FIS (Exógenos)

Uso de los centros de salud públicos (MSSP)
 Talla al nacer
 Peso al nacer
 Duración de la lactancia
 Características del hogar
 Calidad del hogar
 Tipo de hogar
 Accesibilidad a los servicios básicos
 Distancia entre el hogar y el centro de salud más cercano
 Costo de la consulta en el centro de salud
 Gastos del hogar
 Características personales de los miembros del hogar
 Edad
 Idioma
 Nivel de educación
 Ocupación

5. Indicadores de identificación

Costo programado por componente del proyecto
 Gastos reales por componente del proyecto

a. No se considera en los datos básicos.

b. Se elaborará en coordinación con el personal de la Reforma Educacional; se considerará exógena a la intervención, a menos que las intervenciones de la Reforma Educacional del FIS se estén considerando en conjunto.

c. Se ha eliminado la tasa de mortalidad general, tasa de natalidad, tasa de fecundidad global, mortalidad adulta y expectativa de vida al nacer.

d. Se ha eliminado la capacitación en temas de salud.

e. Se ha eliminado la tasa de mortalidad general, tasa de natalidad, tasa de fecundidad global, mortalidad adulta (hombres y mujeres), expectativa de vida al nacer, prevalencia de enfermedades respiratorias agudas y tratamiento de la tos.

Anexo 5

Plantilla de marco lógico para el diseño de un proyecto - Resumen del documento de realización del proyecto o el documento de *Resumen de la evaluación del proyecto*

<i>Jerarquía de objetivos</i>	<i>Indicadores de desempeño clave</i>	<i>Supervisión y evaluación</i>	<i>Supuestos críticos</i>
<p>Meta estratégica de asistencia al país relacionada con el sector: Proporcionar una declaración de una frase sobre la meta estratégica a largo plazo (según se refleja en la CAS) al cual debe contribuir el diseño del proyecto. La declaración debe describir el cambio sustancial en el desarrollo de el o los sectores de interés.</p>	<p>Indicadores del sector:</p> <ol style="list-style-type: none"> Los indicadores correspondientes a la meta de la CAS relativa al sector involucran mediciones que en general no son financiadas por el proyecto, pero que sí pueden ser financiadas por el Banco como parte de otro trabajo. Normalmente el prestamista monitorea estos indicadores como parte de una buena práctica de administración sectorial. 	<p>Informes del sector y país:</p> <ol style="list-style-type: none"> En esta columna se identifica el lugar donde se encontrará la información para verificar cada indicador y el proceso involucrado. Los indicadores correspondientes a la meta de la CAS relativa al sector generalmente se monitorean y/o evalúan a través de varios informes del país o sector generados fuera del proyecto. 	<p>(De la meta a la misión del Banco)</p> <ul style="list-style-type: none"> Suponiendo el logro a largo plazo de la meta de la CAS relativa al sector (declarada en la celda de la extrema izquierda), se enumera cualquier otro supuesto adicional necesario para vincular esta meta a la misión del Banco (esto es, aliviar la pobreza). Estos supuestos a menudo implican condiciones, acciones o respuestas ajenas al proyecto y al sector.
<p>Objetivo de desarrollo del proyecto:</p> <ol style="list-style-type: none"> Proporcionar una declaración de una frase sobre el cambio en el comportamiento que se espera en el grupo o institución(es) destinatarios-beneficiarios al término de la implementación del proyecto. El logro de los objetivos sirve de simple verificación de la demanda por resultados del proyecto. El objetivo debería expresar un solo propósito de 	<p>Indicadores de resultado/impacto:</p> <ol style="list-style-type: none"> Los indicadores de resultado se relacionan con los resultados que deberían lograrse al final de la implementación del proyecto, mientras que el impacto podría no concretarse totalmente hasta cinco o más años después de terminada la implementación del proyecto. Los indicadores en el resultado (nivel de PDO) no son una redeclaración de aquellos a nivel de salida. 	<p>Informes del proyecto:</p> <ol style="list-style-type: none"> En esta columna se identifica el lugar donde se encontrará la información para verificar cada indicador y el proceso involucrado. Los indicadores correspondientes al objetivo de desarrollo del proyecto generalmente se monitorean y/o evalúan a través de varios informes del proyecto, informes de la misión de supervisión e informes de evaluación (intermedio y final). En aquellos casos en que se 	<p>(Del objetivo de desarrollo del proyecto a la meta de CAS del sector)</p> <ul style="list-style-type: none"> Suponiendo que se logre el objetivo de desarrollo del proyecto, se enumera cualquier supuesto adicional necesario para justificar el aporte del proyecto a la meta declarada. Estos supuestos se refieren a la o las contribuciones de proyectos adicionales, aportes adicionales o respuestas adicionales de grupos e

desarrollo que sea realista, específico, mensurable y determinado por la demanda. Una guía para fijar el objetivo del proyecto se puede encontrar en *“Do’s and Dont’s for Setting a Project Development Objective”* (llame a x37065 o envíe un correo electrónico a M&Ehelp@worldbank.org para obtener una copia)

Resultado de cada componente:

1. Declarar aquí (en pasado) el valor agregado por la ejecución de cada componente.
2. Una declaración correcta del valor agregado del resultado será fácil de medir (según lo reflejan los indicadores de la derecha).
3. Para simplificar y aclarar la lógica, debería haber una declaración de resultado por cada componente correspondiente del proyecto.
4. Cada resultado debe corresponder en número a su respectivo componente.
5. El equipo del proyecto generalmente es responsable de asegurar la entrega de los resultados como parte de un

3. La recopilación de datos para medir estos indicadores generalmente es financiada por el proyecto.

Indicadores de resultado:

1. Los indicadores de resultado tienen atributos de cantidad, calidad y tiempo. Si no se establece el tiempo, se asume el término del proyecto.
2. Los indicadores de resultado generalmente incluyen medidas de eficacia en función de los costos.
3. La recopilación de datos para la medición de los indicadores de resultados es financiada por el proyecto.
4. Para proyectos complejos, se puede usar un cuadro separado (tal vez un apéndice al Anexo 1) para proporcionar una lista más detallada de los indicadores de resultados.
5. Es mejor tener sólo algunos pocos indicadores de resultados significativos y fáciles de medir que una abundancia de indicadores para

requiere recopilar datos, es necesario mencionar los métodos y responsabilidades, los que podrían incluir consultas de los beneficiarios.

Informes del proyecto:

1. Los indicadores de salida generalmente se monitorean y/o evalúan a través de varios informes del proyecto, informes de la misión de supervisión e informes de evaluación (intermedio y final).
2. Entre las fuentes de datos para supervisar y evaluar estos indicadores normalmente se incluyen sistemas de registro de gestión y administración e informes de resumen generados por el proyecto.

instituciones beneficiarias que son críticas para el logro de la meta declarada.

(de los resultados al objetivo de desarrollo del proyecto)

- Suponiendo que al final del proyecto se logren los resultados enumerados en la celda de la extrema izquierda, se enumera cualquier supuesto adicional necesario para lograr el objetivo del proyecto.
- Estos supuestos pueden esconder condiciones, cambios de políticas o comportamiento esperados en los grupos o instituciones beneficiarias que son necesarios para el éxito del proyecto.
- Estos supuestos son decisivos para lograr el objetivo declarado del proyecto, pero están más allá del control directo del proyecto.

buen diseño del proyecto y una buena implementación, planificación y entrega.

los cuales la recopilación de datos resulta compleja.
6. Los indicadores de salida se acuerdan con el prestatario durante la etapa de PCD (en cuanto a la disponibilidad de datos y facilidad para recopilarlos) y una línea base que se obtiene antes de la evaluación.

Componentes / subcomponentes del proyecto:

1. Un componente es un grupo de subcomponentes o actividades diseñadas para producir un único resultado del proyecto.
2. Se enumera cada componente del proyecto como un encabezado principal, seguido de los principales subcomponentes, si los hay, que se financian como parte de él.

Aportes del proyecto:

- (presupuesto para cada componente)
1. Se enumeran los aportes del componente en términos del costo total de cada componente, incluyendo contingencias (p. ej., US\$___)
 2. En los proyectos grandes y complejos, también se pueden mostrar los costos de los subcomponentes (indentados, para separarlos de los costos de los componentes).

Informes del proyecto:

1. Los aportes generalmente se supervisan a través de informes de avance e informes de desembolsos (ambos trimestrales).
2. Los aportes generalmente se evalúan a través de informes de misión de supervisión (semestrales) e informes de auditorías (anuales).

(de Componentes del proyecto a Resultados del proyecto)

- Suponiendo que los componentes y las actividades enumeradas en la celda de la extrema izquierda se implementan con éxito, se enumera cualquier otro supuesto necesario para lograr los resultados declarados.
- Estos supuestos son condiciones fuera del control directo del proyecto y son necesarios para lograr los resultados declarados del proyecto.
- El proyecto en sí mismo no debe ser el gastar dinero para lograr alguna de estas condiciones (puesto que tales supuestos se incluyen en los componentes en sí).

Anexo 6


Matriz de análisis

Evaluación del impacto del Fondo de Inversión Social de Emergencia de Nicaragua -- 1998

A. Focalización en la pobreza				
<i>Temas</i>	<i>Indicadores generales</i>	<i>Metodologías</i>	<i>Comentarios</i>	<i>Fuente de datos</i>
Niveles de pobreza de comunidades / distritos del FS	<ul style="list-style-type: none"> • % de hogares en la comunidad o distrito bajo la línea de pobreza o niveles de consumo de la extrema pobreza • Nivel de consumo medio en las comunidades que participan en el fondo social frente al nivel de consumo del país • Índice del mapa de la pobreza (como se usa en el FS) 	<p>Requiere encuesta sobre ingreso y consumo familiar e identificación de actividades del FS por comunidad y distrito</p> <p>Requiere encuesta sobre ingreso y consumo familiar e identificación de actividades del FS por comunidad y distrito</p> <p>Los mapas normalmente usan medidas sustitutivas, como un índice compuesto de la pobreza basado en una combinación de variables</p>	<p>Para comparaciones entre países, se necesitan definiciones de líneas de pobreza similares</p> <p>Sus desventajas son que escoge los indicadores y ponderaciones arbitrariamente y cada país tiene un índice diferente. Su ventaja es que a menudo proporciona más desagregación geográfica que las encuestas de ingreso y consumo– ambos se pueden vincular para derivar niveles predichos de consumo a niveles desagregados.</p>	<p>Sobremuestreo de la encuesta nacional domiciliaria (ENV) en comunidades del FS – sólo para proyectos de educación, salud, agua y saneamiento</p> <p>Sobremuestreo de la encuesta nacional domiciliaria (ENV) en comunidades del FS – sólo para proyectos de educación, salud, agua y saneamiento</p> <p>El FS usa un mapa de la pobreza basado en los datos de la ENV93, utilizando un índice compuesto de la pobreza; se actualizará con la ayuda de la ENV98 y datos del Censo para predecir el consumo a niveles subnacionales.</p>

Niveles de pobreza de beneficiarios del FS (a nivel de hogar)	<ul style="list-style-type: none"> • Porcentaje de beneficiarios bajo la línea de pobreza o en condiciones de extrema pobreza • Nivel medio de consumo de los hogares beneficiarios frente al promedio nacional de hogares similares por tipo de proyecto (p. ej. con niños en la escuela primaria, con acceso a agua por cañería, que usan letrinas) 	<p>Encuesta de ingresos/consumo que capta a los beneficiarios del FS ya sea por el tamaño del FS o por un sobremuestreo en las comunidades del FS</p> <p>Encuesta de ingreso/consumo que capta a los beneficiarios del FS ya sea por el tamaño de la muestra del FS o por el sobremuestreo en las comunidades del FS. También se puede sobremuestrear en las comunidades “pareadas” sin intervenciones del FS.</p>	Puede variar ampliamente por tipo de proyecto del FS	<p>Sobremuestreo de la encuesta nacional domiciliaria (ENV) en comunidades del FS</p> <p>Sobremuestreo de encuesta nacional domiciliaria (ENV) en comunidades del FS</p>
Distribución de recursos del FS	<ul style="list-style-type: none"> • Porcentaje de proyectos/recursos del FS en el quintil inferior de los distritos 	Se necesita una metodología de clasificación coherente entre los países		Se necesita revisar el sistema de clasificación y redimensionarlo

(La matriz continúa en la página siguiente)

Características del diseño institucional que afectan el desempeño de la focalización del FS	<ul style="list-style-type: none"> • Uso del mapa de pobreza • Esfuerzos de promoción • Acceso directo por los grupos beneficiarios • Proporción de proyectos por organismo solicitante • Oficinas descentralizadas • Asignaciones de recursos previstos • Menú de subproyecto 	Crear variables institucionales estándares que se puedan usar para explicar los resultados de la focalización — variables que se puedan obtener fácilmente de los FS	Información disponible del FS 	
Otros factores que afectan el desempeño de la focalización	<ul style="list-style-type: none"> • Antigüedad del FS • ‘Capital social’ de la comunidad • Distancia a la oficina central del FS • El mayor nivel de educación de los beneficiarios • Presencia de intervenciones del gobierno o de ONG • Grado de desigualdad en el ingreso del país 	También se necesitan definiciones estándares de las variables – variables obtenidas del FS, encuestas domiciliarias (con identificación de beneficiarios del FS) y encuestas nacionales	El único indicador dudoso es el ‘capital social de la comunidad’	
Comparación de alternativas	<ul style="list-style-type: none"> • Porcentaje de proyectos y recursos del FS en el quintil inferior de los distritos frente a otros programas y mecanismos de entrega comparables 	Comparar el desempeño de la focalización basada en la ubicación geográfica o niveles de pobreza de los beneficiarios, dependiendo del diseño de la encuesta, escala de FS y otros programas	Dificultad para encontrar comparadores viables; se necesita recopilación de información separada de otros programas comparadores	Planificado para el análisis de eficacia en función de los costos

B. Beneficios				
<i>Temas</i>	<i>Indicadores generales</i>	<i>Fuentes de datos/metodologías</i>	<i>Comentarios</i>	<i>Estudio de caso: Nicaragua</i>
Capital físico	<ul style="list-style-type: none"> Medida en la cual los subproyectos responden a las prioridades de la comunidad Percepción de los beneficiarios sobre el nivel de los beneficios y mejoramiento del bienestar Mejoramiento en el acceso a infraestructura social y económica (antes y después) 	Encuesta a nivel de comunidad, evaluación de los beneficiarios o encuesta domiciliaria con sobremuestras en áreas del FS Encuesta domiciliaria o evaluación de los beneficiarios en comunidades del FS Encuesta domiciliaria de beneficiarios del FS.	Se necesita tener datos básicos o preguntas recordatorias. Se necesita crear indicadores separados por tipo de proyecto de FS	Cubierto en la Evaluación de beneficiarios y la Encuesta de instalaciones financiada por el BID
	<ul style="list-style-type: none"> Mejoramiento en el acceso a infraestructura social y económica frente a proyectos comparadores 	Encuesta domiciliaria de FS y beneficiarios de proyectos comparadores.	Se necesita contar con datos básicos o preguntas recordatorias. Se necesitan crear indicadores separados por tipo de proyecto de FS	Cubierto en la encuesta domiciliaria y en la Evaluación de beneficiarios financiada por el BID Algunas preguntas recordatorias en la encuesta domiciliaria. También posible ex-ante de ENV anteriores; se pueden comparar los beneficiarios del FS con las características de la población nacional y con comunidades pareadas. Se pueden comparar los beneficiarios del FS con las características de la población nacional y con comunidades pareadas.
	<ul style="list-style-type: none"> Mejoramiento en la calidad de la infraestructura y servicios (antes y después) 	Encuesta de instalaciones y encuesta domiciliaria, alguna cobertura de la evaluación de los beneficiarios	Se necesita contar con datos básicos o preguntas que impliquen recordar. Se necesita crear indicadores separados por tipo de proyecto de FS	Para educación, salud, agua y servicios sanitarios, preguntas recordatorias e información histórica de la encuesta de instalaciones y evaluaciones ex-ante del FS
	<ul style="list-style-type: none"> Mejoramiento en la calidad de la infraestructura y servicios frente a proyectos comparadores 	Encuesta de instalaciones y encuesta domiciliaria, alguna cobertura de la evaluación de los beneficiarios (en FS y comparadores)	Se necesita contar con datos básicos o preguntas recordatorias. Se necesita crear indicadores separados por tipo de proyecto de FS	Para educación, salud, agua y servicios sanitarios, instalaciones del FS y no del FS a través de encuesta domiciliaria y de instalaciones

Capital humano	<ul style="list-style-type: none"> • Mejor situación educacional: asistencia a la escuela, años aprobados, índices de repitencia y deserción (antes y después y frente a comparadores) 	Encuesta domiciliaria e información de la escuela		Encuesta domiciliaria e información de la escuela de escuelas y hogares con FS y sin FS
	<ul style="list-style-type: none"> • Mejor estado de salud: incidencia de enfermedades, mortalidad infantil, desnutrición, aumento en el amamantamiento, etc. (antes y después y frente a comparadores) 	Encuesta domiciliaria con módulo de salud. Medidas antropométricas si se incluye la desnutrición		Encuesta domiciliaria e información del centro de salud, para centros de salud y hogares con FS y sin FS
	<ul style="list-style-type: none"> • Mejor situación económica - mayor ingreso, menos tiempo para obtener agua, menor costo de los servicios, mayor empleo (antes y después y frente a comparadores) 	Encuesta domiciliaria		Encuesta domiciliaria para comunidades pareadas beneficiarias y no beneficiarias del FS
Capital social	<ul style="list-style-type: none"> • Mayor capacidad de la comunidad para enfrentar problemas (frente a comparadores) 	Encuesta domiciliaria, encuesta comunitaria y/o evaluación de los beneficiarios		No se aborda
	<ul style="list-style-type: none"> • Mayores índices de participación en cambios impulsados por la comunidad (frente a comparadores) 	Encuesta domiciliaria, encuesta comunitaria y/o evaluación de los beneficiarios	Es necesario crear indicadores	Información en la encuesta domiciliaria sobre participación.

(La matriz continúa en la página siguiente)

C. Sostenibilidad de los beneficios				
<i>Temas</i>	<i>Indicadores generales</i>	<i>Fuentes de datos/metodologías</i>	<i>Comentarios</i>	<i>Estudio de caso: Nicaragua</i>
Sostenibilidad de las operaciones	<ul style="list-style-type: none"> • Condiciones bajo las cuales están operando los proyectos del FS después de la intervención del FS (sostenibilidad absoluta) 	Encuesta de instalaciones	Se puede obtener cierta información adicional de la evaluación de los beneficiarios	Para proyectos de educación y salud, ambos con FS y sin FS
	<ul style="list-style-type: none"> • Condiciones bajo las cuales están operando los proyectos del FS después de la intervención del FS frente a proyectos comparadores (en relación a la sostenibilidad) 	Encuesta de instalaciones	Se puede obtener cierta información adicional de la evaluación de los beneficiarios	Para proyectos de educación y salud, ambos con FS y sin FS
Sostenibilidad del mantenimiento	<ul style="list-style-type: none"> • Mantenimiento de la infraestructura y servicios en el tiempo (absoluta) 	Encuesta de instalaciones	Se puede obtener cierta información adicional de la evaluación de los beneficiarios	Para proyectos de educación y salud, ambos con FS y sin FS
	<ul style="list-style-type: none"> • Mantenimiento de la infraestructura y servicios en el tiempo frente a proyectos comparadores (relativo) 	Encuesta de instalaciones	Se puede obtener cierta información adicional de la evaluación de los beneficiarios	Para proyectos de educación y salud, ambos con FS y sin FS
Sostenibilidad del impacto	<ul style="list-style-type: none"> • Calidad y cantidad de infraestructura y servicios en el tiempo 	Encuesta de instalaciones y encuesta domiciliaria	Se puede obtener cierta información adicional de la evaluación de los beneficiarios	Para proyectos de educación y salud, ambos con FS y sin FS
Sostenibilidad de los impactos en la comunidad	<ul style="list-style-type: none"> • Tendencia de las comunidades con FS a presentar otras propuestas (al FS y ante otros) en el tiempo 	Base de datos del FS, encuesta comunitaria y/o evaluación de los beneficiarios		Se necesita trabajo adicional
	<ul style="list-style-type: none"> • Participación de la comunidad en las necesidades de infraestructura social y económica en el tiempo 	Encuesta comunitaria, encuesta domiciliaria y/o evaluación de los beneficiarios		Se incluye en la <i>siguiente</i> evaluación de los beneficiarios; evaluación del impacto

D. Aspectos de eficacia en función de los costos

<i>Temas</i>	<i>Indicadores generales</i>	<i>Fuentes de datos/metodologías</i>	<i>Comentarios</i>	<i>Estudio de caso: Nicaragua</i>
Eficacia en función de los costos de los subproyectos	<ul style="list-style-type: none"> Costo promedio por escuela, posta de salud, sistema de agua nuevos, frente a enfoques alternativos y proyectos comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales	El cambio en los costos en el tiempo y los proyectos comparadores deben ser idénticos	Datos del FS y no del FS de la encuesta de instalaciones. Las estimaciones de costos no del FS pueden no ser confiables
	<ul style="list-style-type: none"> Costos unitarios: costo por metro cuadrado de construcción, por kilómetro de camino, etc., frente a proyectos comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales		Se pueden calcular los promedios del FS. Se incluirán en el análisis de eficacia en función de costos
	<ul style="list-style-type: none"> Costo promedio por beneficiario por tipo de proyecto del FS frente a los comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales		Se pueden calcular los promedios del FS. Se incluirán en el análisis de eficacia en función de costos
	<ul style="list-style-type: none"> Costo promedio del empleo generado frente a los comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales		Se pueden calcular los promedios del FS. Se incluirán en el análisis de eficacia en función de costos
Eficacia en función de los costos de los mecanismos de entrega	<ul style="list-style-type: none"> Costos institucionales del FS (de inversión y operativos) como proporción de los proyectos del FS frente a proyectos comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales	Necesidad de crear definiciones estándares de los costos institucionales; especificar período	Se pueden calcular los promedios del FS. Se incluirán en el análisis de eficacia en función de costos
	<ul style="list-style-type: none"> Tiempo de ejecución promedio frente a proyectos comparadores 	Base de datos del FS e información de ministerios estatales y gobiernos municipales		Se pueden calcular los promedios del FS. Se incluirán en el análisis de eficacia en función de costos