

## 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<br>No califica<br>[b=1;e=0;c=0]                    |                                     | Hogares con más de<br>½ acre  | B<br>No calificaría<br>[b=0;e=0;c=0] |
| C<br>Califica, pero no<br>participa<br>[b=1;e=1;c=0] | D<br>Participantes<br>[b=1;e=1;c=1] |                               | E<br>Calificaría<br>[b=0;e=1;c=0]    |
|  |                                     | Hogares con ½ acre y<br>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.

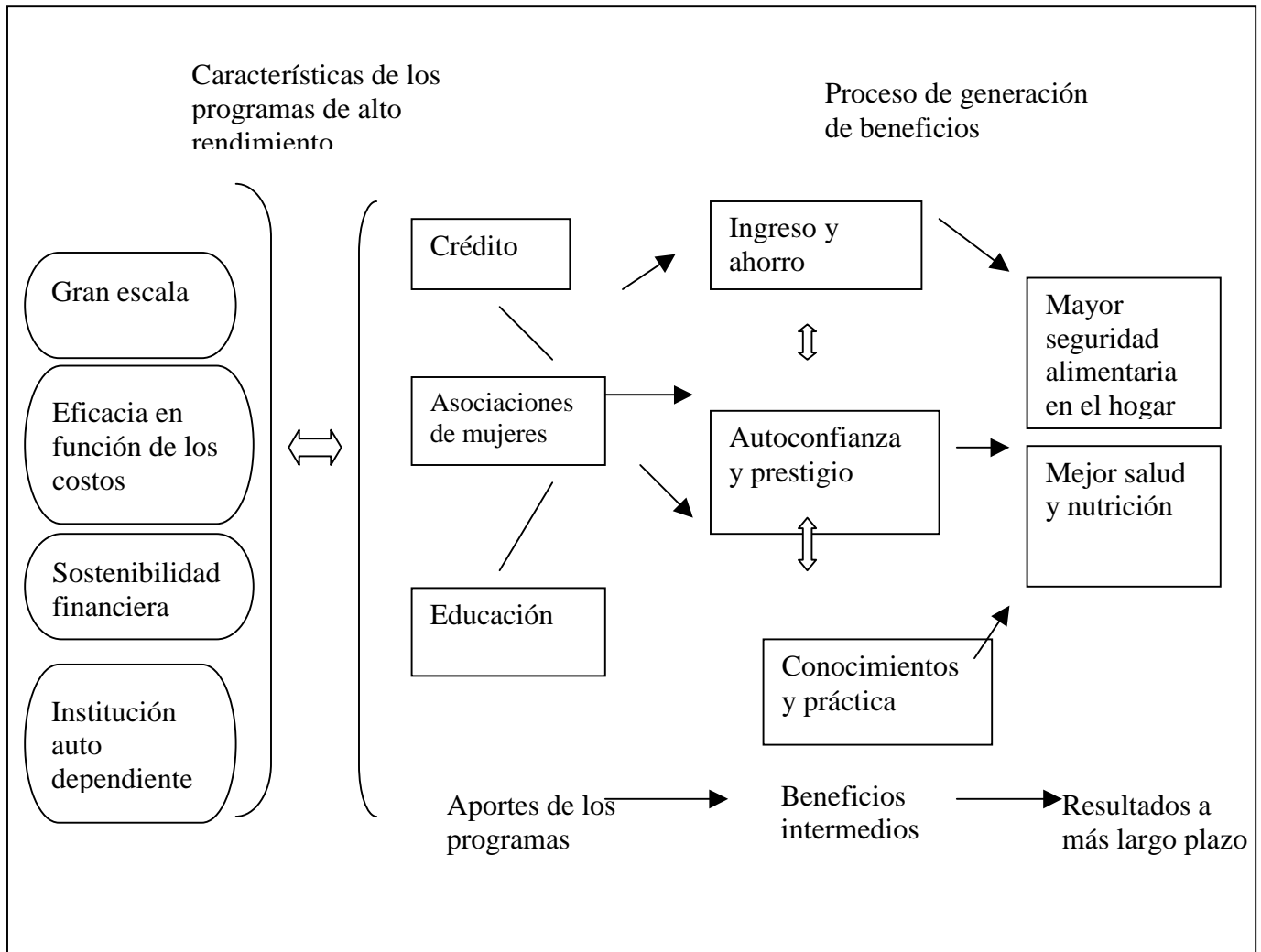
## VIII. Fuentes

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  $\mu$  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  $\Delta^*$ . 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 ( $\beta_{1j}$  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  $\delta_1$  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.