

EVALUACIÓN EN UN MUNDO REAL

TRABAJANDO BAJO RESTRICCIONES
PRESUPUESTARIAS, CRONOLÓGICAS, DE
INFORMACIÓN Y POLÍTICAS

VISIÓN GENERAL

MICHAEL BAMBERGER
JIM RUGH
LINDA MABRY

Junio 2007

Adaptado del Capítulo 16, “Bringing it all together” (“Articulándolo todo”), de “RealWorld Evaluation: Working under Budget, Time, Data and Political Constraints”, por Michael Bamberger, Jim Rugh y Linda Mabry, Sage Publications, 2006.

Los autores agradecen a UNICEF por su gentileza
en financiar la version Espanola del documento

Índice

		Pág.
1	Delimitación del alcance de la evaluación <ul style="list-style-type: none"> ▪ Determine las necesidades del cliente ▪ Determine el entorno político ▪ Defina la teoría del programa ▪ Adapte los planes de evaluación 	4
2	Selección del mejor diseño entre las opciones disponibles <ul style="list-style-type: none"> ▪ Afine el diseño de evaluación ▪ Identifique los tipos de análisis y comparaciones que resultan cruciales para la evaluación ▪ Evalúe las amenazas a la validez e idoneidad de diferentes Diseños 	15
3	Determinación de la metodología apropiada	24
4	Opciones para fortalecer los diseños de evaluación en un mundo real <ul style="list-style-type: none"> ▪ Sustente el diseño de evaluación en el modelo teórico del programa ▪ Complemente la evaluación cuantitativa (sumativa) con una evaluación de procesos ▪ Incorpore el análisis contextual ▪ Reconstruya la situación de base ▪ Use enfoques de métodos mixtos para fortalecer la validez de los indicadores y mejorar la interpretación de los hallazgos ▪ Controle las diferencias entre los grupos del proyecto y de comparación 	25
5	Economizar en recursos humanos para la evaluación <ul style="list-style-type: none"> ▪ Utilice consultores/as externos con inteligencia ▪ Considere el uso de especialistas en áreas de contenido ▪ Sea creativo en relación con los recolectores/as de datos 	32
6	Recolectar los datos con eficiencia <ul style="list-style-type: none"> ▪ Simplifique los planes de recolección de datos ▪ Encargue estudios preparatorios ▪ Busque datos secundarios confiables ▪ Recolecte solo los datos necesarios ▪ Busque formas simples de recolectar información sobre temas sensibles y de poblaciones difíciles de acceder 	36
7	Analizar los datos con eficiencia <ul style="list-style-type: none"> ▪ Busque formas de manejar los datos con eficiencia ▪ Concentre el análisis en responder las preguntas clave 	39
8	Informar sobre los hallazgos de manera eficiente y efectiva <ul style="list-style-type: none"> ▪ Informe sucinto a los clientes primarios ▪ Informes prácticos, entendibles y útiles para otros destinatarios 	43
9	Ayudar a los clientes a utilizar los hallazgos correctamente	47

Figuras

1	El enfoque de evaluación en un mundo real	6
2	Modelo teórico simple de un programa	13

Tablas [*al final del texto*]

1	Algunas de las formas en que las influencias políticas afectan las evaluaciones	49
2	Cinco estrategias de evaluación y los diseños correspondientes	52
3	Los siete diseños de evaluación de impacto más ampliamente utilizados	54
4	Fortalezas y debilidades de los siete diseños cuasi-experimentales más frecuentemente utilizados	56
5	Reducción de los costos de recolección y análisis de datos para evaluaciones cuantitativas y cualitativas	58
6	Ahorro de costos estimado en el caso de EMR menos robustas en comparación con el Diseño 2	61
7	Factores que afectan el tamaño de la muestra	62
8	Reducción del tiempo requerido para la recolección y el análisis de datos en evaluaciones cuantitativas y cualitativas	64
9	Métodos de recolección rápida de datos	67
10	Estrategias para responder a las restricciones de información y reconstruir la situación de base	71
11	Factores que afectan la idoneidad del diseño y los hallazgos de la evaluación	75
12	Algunas amenazas que afectan la validez incluso de los diseños cuantitativos más robustos	76
13	Características de los enfoques cuantitativos y cualitativos en diferentes etapas del proceso de evaluación	78
14	Elementos de un enfoque de investigación multidisciplinario integral	82

Anexos

1	Lista de verificación estandarizada para evaluar la idoneidad y la validez de todos los diseños de evaluación	84
2	Lista de verificación para evaluar las amenazas a la validez de los diseños de evaluación cuantitativos	87

Referencias		91
-------------	--	----

ARTICULÁNDOLO TODO

Aplicación de enfoques de Evaluación en un Mundo Real en cada etapa del proceso de evaluación

Introducción a esta versión independiente

En este capítulo del libro analizamos cómo los enfoques de Evaluación en un Mundo Real (**EMR**) pueden ser aplicados en cada etapa del diseño y la ejecución de una evaluación típica. Identificamos los problemas que pueden surgir para una EMR – es decir, ahí donde existen restricciones de financiamiento, tiempo, disponibilidad de información y concepciones preestablecidas de los clientes – y sugerimos formas en que el enfoque de EMR puede ayudar a superar dichas limitaciones. Los lectores y lectoras neófitos en el campo de la evaluación podrían beneficiarse también de este capítulo como una introducción general a la planificación, diseño, ejecución, difusión y uso de una evaluación.

El presente capítulo está diseñado como una introducción a la EMR y condensa al mismo tiempo muchos de los principales aspectos tratados en el libro. La Figura 1 resume los siete pasos del enfoque de EMR e incluye además referencias a otros capítulos del libro, donde encontrarán un tratamiento más detallado de los temas individuales¹.

Algunas de las tablas importantes de otros capítulos han sido incluidas al final de esta versión independiente del Capítulo 16.

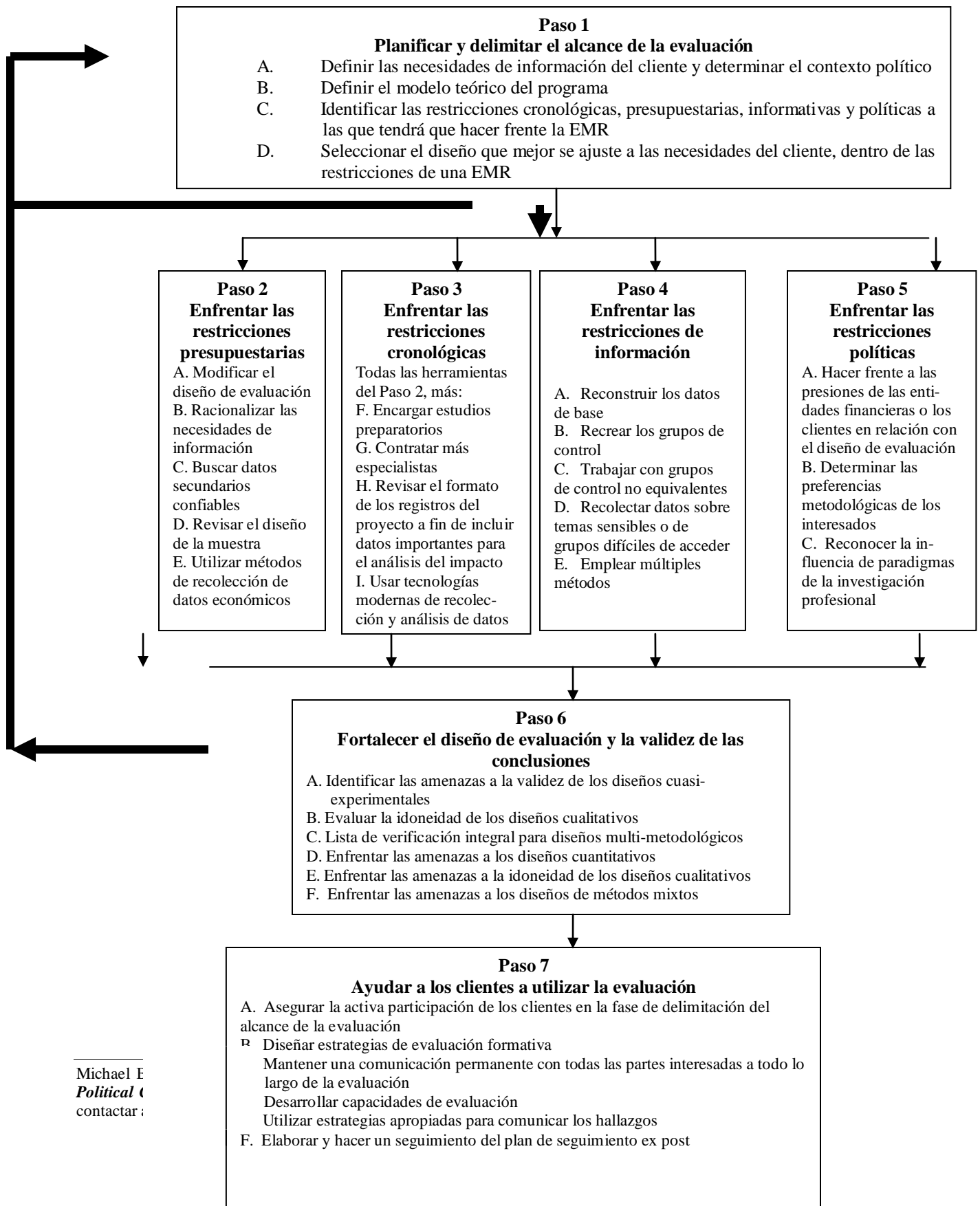
1. Delimitación del alcance de la evaluación

Es importante que los responsables de llevar a cabo una evaluación establezcan claramente qué es lo que esperan las personas que están pidiendo una evaluación (los **clientes**² y los **interesados**) – esto es, el contexto político dentro del cual se ejecutarán tanto el **proyecto** como la evaluación. Del mismo modo, es importante conocer las decisiones operativas y de políticas a las que ha de contribuir la evaluación, así como el nivel de **precisión** requerido para proveer la información en la que se sustentarán dichas decisiones (ver el Capítulo 2).

¹ Las referencias a los capítulos aluden al libro *RealWorld Evaluation* (“Evaluación en un Mundo Real”).

² Ver definiciones de los términos **en negrita** en el Glosario del libro.

Figura 1. El enfoque de Evaluación en un Mundo Real



Michael E
Political &
contactar :

Determine las necesidades del cliente.

Un primer paso esencial de la preparación para una evaluación es establecer claramente las prioridades y necesidades de información del cliente (la organización u organizaciones que encargan la evaluación) así como de los otros interesados (las personas interesadas en o afectadas por el proyecto). El plazo, la focalización y el nivel de detalle de la evaluación dependerán de las necesidades de información del cliente y de los tipos de decisiones a los que ha de contribuir la evaluación.

El proceso de establecer qué preguntas deben ser respondidas puede ayudar a las personas que están planeando una evaluación a identificar formas de eliminar la recolección y el análisis de datos innecesarios, logrando con ello una reducción de los costos y plazos. El evaluador/a en un mundo real debe distinguir entre (a) información que resulta esencial para responder a las preguntas clave que sostienen la evaluación, y (b) preguntas adicionales que sería interesante formular si se dispusiera del tiempo y los recursos necesarios, pero que posiblemente deban ser omitidas dadas las limitaciones a las que está sujeta la evaluación.

Una decisión adicional relacionada con los costos y plazos puede ser quién debe participar en la recolección de datos y en la revisión de los informes de evaluación. Muchos proyectos de desarrollo se rigen por la filosofía de promover el empoderamiento de los miembros de la comunidad y otros interesados (por ejemplo, administradores escolares y maestros/as, personal de los centros de salud, instituciones estatales locales y organizaciones no gubernamentales), lo cual incluye invitarlos a participar en las actividades de seguimiento y evaluación. Una pregunta que podría surgir a la luz de las restricciones que enfrenta una EMR es: “¿Cuán importante es incluir métodos participativos y una representación adecuada de los participantes del proyecto y otros interesados en la realización de la evaluación?” Los métodos participativos de recolección de datos tienden a ser más caros y requieren plazos más largos, porque se tiene que programar el tiempo suficiente para desarrollar una buena relación con la comunidad y otros interesados y edificar una relación de confianza. El evaluador en un mundo real, consciente de los costos, tiene que determinar si el cliente y los interesados centrales atribuyen a los enfoques participativos el suficiente valor como para asignarles el tiempo y el presupuesto requeridos *para hacerlos bien*.

Otro reto para la EMR es que a menudo resulta más caro y requiere plazos más largos llegar a los grupos más pobres y vulnerables; por eso, cuando se está sujeto a restricciones cronológicas y presupuestarias a menudo se enfrentarán presiones para excluir a dichos grupos de las consultas. “Sería realmente magnífico consultar con los invasores que no tienen títulos de propiedad sobre los terrenos, pero lamentablemente... simplemente no tenemos ni el dinero ni el tiempo.”

Una función importante de la etapa de delimitación del alcance de la evaluación es averiguar si la omisión de consultar con los grupos afectados por el proyecto, entre ellos los sectores más pobres y vulnerables, se debe a la falta de recursos o a la baja prioridad que el cliente otorga a la participación de dichos grupos. Con frecuencia se recurre a la falta de tiempo y de dinero como excusa. Por eso, es importante que el evaluador/a establezca con precisión cuál es la perspectiva del cliente antes de tomar la decisión relativa al enfoque que debe ser adoptado.

Determine el entorno político.

El entorno político incluye: las prioridades y perspectivas del cliente y otros interesados clave, la dinámica de poder y las relaciones entre los primeros y actores clave del proyecto que está siendo evaluado, e incluso las tendencias o preferencias filosóficas o metodológicas de las personas que efectúan la evaluación. La Tabla 1 contiene algunas de las formas en que factores políticos pueden afectar una evaluación en la etapa de diseño, mientras están siendo ejecutadas y cuando se están presentando y difundiendo los hallazgos (ver también el Capítulo 6 al respecto).

Es importante evitar el supuesto de que toda influencia política es mala y que los evaluadores/as deben poder llevar a cabo la evaluación de la manera que consideren más idónea, sin interferencias de políticos u otros interesados “con estrechez de miras”, que intentan cerciorarse de que se tomen en cuenta sus intereses en la evaluación. El propósito medular de la evaluación es contribuir a echar luces sobre políticas y **programas** respecto de los cuales la gente tiene opiniones fuertes, y a menudo contradictorias. Si una evaluación no es objeto de presiones o influencias políticas de ningún tipo, probablemente signifique ya sea que el tema de estudio no posee ninguna pertinencia para nadie o que la evaluación está diseñada de tal modo que a los grupos concernidos no les es posible expresar sus puntos de vista. Los evaluadores/as nunca deben asumir que son ellos quienes tienen la razón y que los interesados que tienen opiniones diferentes a las suyas sobre temas centrales, la metodología apropiada o la interpretación de los **hallazgos**, están prejuiciados, mal informados o simplemente equivocados. En la **Lista de verificación estandarizada para evaluar la idoneidad y la validez de los diseños de evaluación** (Anexo 1), el criterio de evaluación C tiene que ver con la **validez interna** y la **autenticidad** de los hallazgos de la evaluación: “¿Son creíbles los hallazgos para las personas que están siendo evaluadas y para los lectores/as, y ofrecen un retrato veraz de lo que se está estudiando?”.

Si grupos clave consideran que el análisis no es verosímil, en ese caso el evaluador/a tendrá que retroceder y revisar cuidadosamente la metodología y los supuestos subyacentes. Suspirar y pensar qué difícil es lograr que el cliente “entienda” los hallazgos y recomendaciones no constituye nunca una respuesta apropiada.

Una de las dimensiones del análisis contextual que se emplea para la elaboración del modelo teórico del programa (ver sección siguiente) es examinar la influencia de los factores políticos. Muchas de las dimensiones contextuales - económicas, institucionales, ambientales y socioculturales – influyen en la forma en que grupos políticamente interesados verán el proyecto así como la evaluación del mismo. La comprensión cabal de dichos factores contextuales es fundamental para entender las actitudes de interesados clave frente al programa y a la evaluación del mismo. Tener una idea clara de dichos intereses ayuda a identificar formas de hacer frente a las presiones que ejercen dichos interesados sobre la evaluación.

No sorprende que muchas evaluaciones de programas sean encargadas con motivaciones políticas en mente. Un cliente puede planear emplear una evaluación para generar apoyo para su programa y por lo tanto puede poner resistencia a la inclusión de hallazgos que no sean sino positivos. Por otro lado, la intención real pero no divulgada que puede haber tenido un cliente para encargar una evaluación puede ser reunir los argumentos necesarios para despedir a un/a gerente o cerrar un proyecto o departamento.

Dichas razones rara vez se transmiten de forma explícita. Del mismo modo, distintos interesados pueden tener opiniones fuertemente divergentes acerca de un programa, su ejecución, sus motivaciones, sus líderes y cómo debe ser evaluado. Las personas que se oponen a una evaluación pueden ser capaces de prevenir la evaluación u obstruir el acceso a datos, la aceptación de los resultados de la evaluación o la continuación de un contrato de evaluación.

Antes de iniciar una evaluación, el evaluador/a debe anticipar los diferentes tipos de consideraciones políticas que pueden afectar la evaluación e intentar explorarlas, ya sea directa o indirectamente, junto con el cliente e interesados clave. El Capítulo 6, Tabla 6.1^{3*}, ilustra algunas de las numerosas formas en que el contexto político puede afectar la forma en que se diseña, ejecuta, difunde y utiliza una evaluación.

Las dimensiones políticas incluyen no solamente a los clientes y otros interesados. Incluyen también la preferencia de los evaluadores/as individuales por enfoques que calzan con sus opiniones personales y profesionales acerca de lo que constituye una práctica competente y apropiada. Diferentes evaluadores/as – incluso aquellos que han elegido trabajar juntos en un proyecto – pueden tener posiciones distintas respecto de sus responsabilidades públicas y éticas. Los evaluadores/as, como todas las personas, tienen sus propios valores personales. Sin embargo, muchos evaluadores/as posiblemente se sientan más cómodos si consideran el trabajo de evaluación no como una imposición de los valores del evaluador/a sino, más bien, como un juicio imparcial y objetivo, sustentado en datos, sobre los méritos, las deficiencias, la eficacia, la eficiencia y el logro de las metas de un programa. Los evaluadores/as deben ser conscientes de sus propias perspectivas (y prejuicios) e intentar cerciorarse de que éstas sean reconocidas y tomadas en cuenta en la evaluación. (Ver la Sección A de la “Lista de verificación estandarizada para evaluar la idoneidad y la validez de los diseños” en el Anexo 1.)

Los clientes pueden sustentar la selección de los evaluadores/as en la reputación de éstos de honestidad sin compromisos, apoyándose en dicha reputación para garantizar la **credibilidad** y la aceptación de los hallazgos. O la selección del evaluador/a puede obedecer a las posturas ideológicas adoptadas por el evaluador/a y que coinciden con las del cliente. Estas decisiones pueden ser tan poco explícitas que inicialmente pueden pasar desapercibidas en las negociaciones amigables y afirmaciones entusiastas sobre la importancia estratégica de la evaluación propuesta. El Capítulo 6 versa sobre algunas de las opciones disponibles para el evaluador/a cuando éste considera algunas de las presiones de los clientes como ética o profesionalmente inaceptables.

Los evaluadores/as deben estar alertas asimismo frente al hecho de que la orientación política de los clientes y otros interesados puede influir en la forma en que se difunden y utilizan los hallazgos de una evaluación. En ocasiones los clientes pueden ignorar los hallazgos que no les agradan y prevenir su distribución, circulando los informes únicamente entre lectores/as cuidadosamente seleccionados, difundiendo solamente resúmenes abreviados y suavizados, y asumiendo la responsabilidad de presentar los informes a las juntas directivas o instituciones financieras y luego actuar en

³ Las tablas de *RealWorld Evaluation* que están marcadas con un asterisco (*) no han sido incluidas en esta visión general.

esa capacidad de manera manipuladora. Son conocidos los casos de clientes que realizan presentaciones orales e incluso ofrecen testimonios distorsionando los hallazgos de una evaluación, emprenden actividades de seguimiento ex post no sugeridas y hasta contraindicadas por los informes de evaluación y desacreditan evaluaciones y a evaluadores/as que representan una amenaza para sus programas y su prestigio.

El evaluador/a inteligente debe ser consciente de dichas realidades y estar preparado para lidiar con ellas de manera apropiada. El Capítulo 6 sugiere algunas estrategias de las que dispone el evaluador en un mundo real para hacer frente a las restricciones políticas arriba indicadas y otras, tanto durante el diseño de la evaluación como durante su ejecución y la presentación y el uso de los hallazgos.

Defina la teoría del programa.

Antes de realizar una evaluación, es necesario identificar el modelo teórico o lógico, ya sea explícito o implícito, que subyace al diseño sobre el cual descansa un proyecto (ver el Capítulo 9). Una función importante de la evaluación del impacto es validar la hipótesis de que las intervenciones y los productos del proyecto han contribuido a los efectos directos deseados, los cuales, junto con los factores externos que el proyecto asumió que prevalecerían, deberían generar un impacto sostenible.

Definir el modelo teórico o lógico de un programa representa una buena práctica para cualquier evaluación, pero es especialmente útil en el caso de una EMR, en la que, debido a restricciones presupuestarias, cronológicas y de otra índole, es necesario priorizar los aspectos en los que debe concentrarse la evaluación. Una revisión inicial de lo que el proyecto hizo con base en su modelo lógico podría revelar datos faltantes o información requerida para verificar la solidez del razonamiento lógico y si el proyecto pudo llevar adelante las acciones que se requerían para lograr el impacto deseado.

Si el modelo lógico se encuentra claramente expresado en el plan del proyecto, puede ser utilizado para orientar la evaluación. Si éste no es el caso, el evaluador/a tendrá que construirlo a partir de una revisión de la documentación del proyecto y a través de discusiones con la entidad ejecutora del proyecto, los participantes del proyecto y otros interesados (ver el Capítulo 9). En muchos casos, esto supone un proceso iterativo en el que el diseño del modelo lógico emerge a medida que vamos conociendo más detalles sobre la evaluación, en el transcurso de la misma.

Además de expresar la teoría interna causa-efecto a partir de la cual se diseñó el proyecto, el modelo lógico debería identificar además las características socioeconómicas de los grupos poblacionales afectados, así como factores contextuales tales como las condiciones económicas, políticas, institucionales, psicológicas y ambientales que afectan a la comunidad objetivo.

Las fases o niveles centrales de un modelo lógico simple pueden resumirse de la siguiente manera (ver la Figura 16.1* y el Capítulo 2):

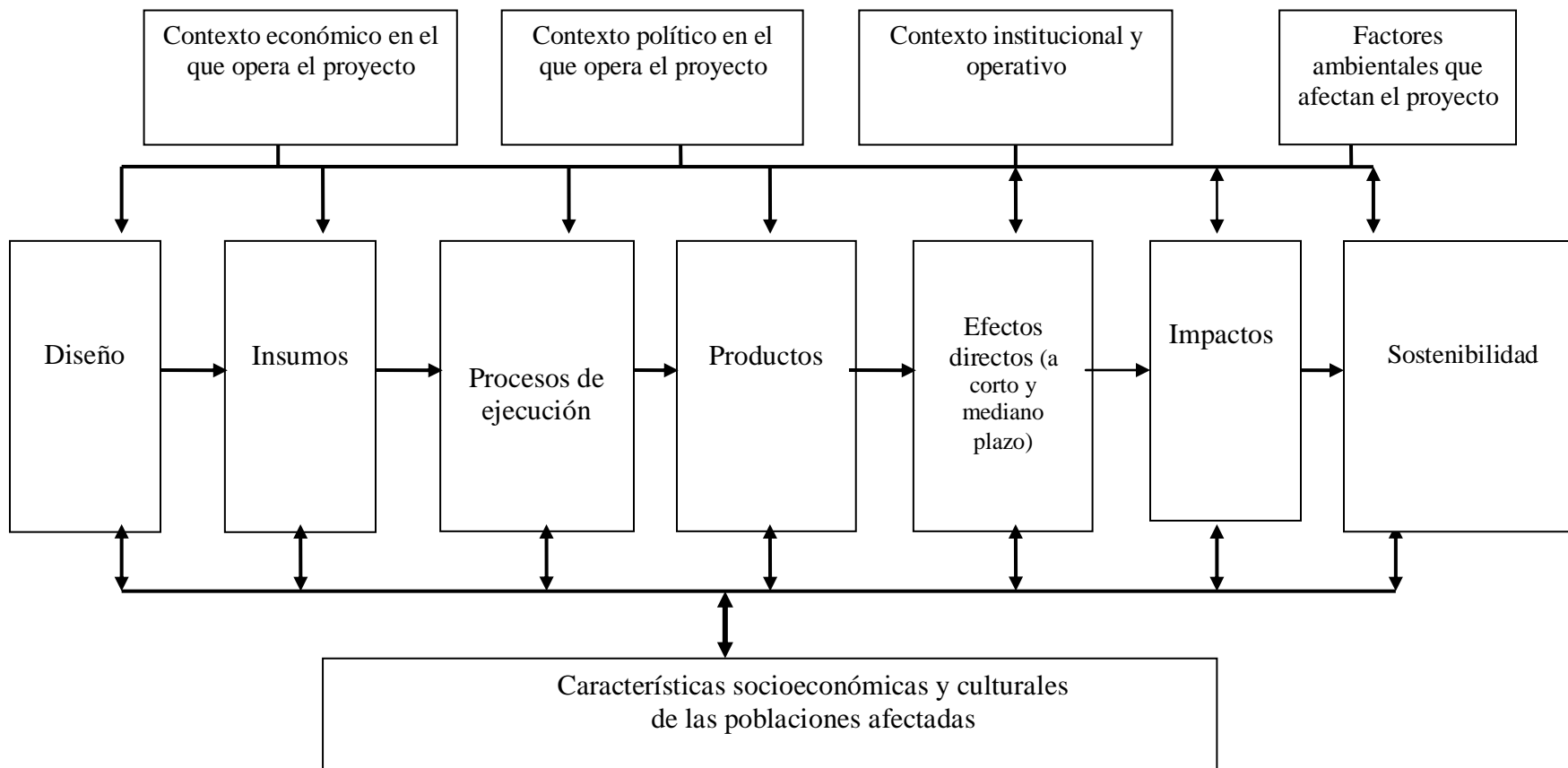
1. **Diseño.** ¿De qué manera se diseñó el proyecto? ¿Quién diseñó este proyecto? ¿Fue diseñado solamente por un grupo reducido del personal de la institución financiera o ejecutora o por un consultor/a externo? ¿O hubo una extensa participación de una mezcla de interesados, entre ellos los beneficiarios previstos? ¿Se basó el diseño en un diagnóstico holístico de la situación de las comunidades objetivo? ¿Y se

consideraron en el diseño las enseñanzas aprendidas de evaluaciones de proyectos anteriores que utilizaron enfoques similares en condiciones semejantes?

2. **Insumos**⁴. Los insumos son los recursos financieros, humanos, materiales, tecnológicos e informativos que han sido empleados en una intervención para el desarrollo.
3. **Proceso de ejecución**. Incluye las acciones emprendidas o el trabajo realizado a través de los cuales se movilizan insumos tales como fondos, asistencia técnica y otros tipos de recursos con el fin de generar productos específicos. Uno de los factores principales aquí es si los beneficiarios previstos y otros interesados participaron en el proceso de ejecución y cómo.
4. **Productos**. Los productos son los bienes de capital, productos y servicios que resultan de la ejecución de las actividades de una intervención para el desarrollo. Tenga presente que los ejecutores del proyecto ejercen un control directo sobre los productos – aunque no sobre los factores contextuales externos que pueden afectar la ejecución oportuna o la calidad de los productos.
5. **Efectos directos**. Se trata de los resultados a corto y mediano plazo probables o logrados por los productos de una intervención. Tenga presente que los ejecutores del proyecto no ejercen un control directo sobre los efectos. Los efectos directos son lo que otros hacen por su cuenta, pero influenciados por los productos del proyecto.
6. **Impactos**. Impactos son los efectos a largo plazo, positivos y negativos, producidos directa o indirectamente por una intervención para el desarrollo, intencionalmente o no, en grupos poblacionales identificables. Los impactos pueden ser económicos, socioculturales, institucionales, ambientales, tecnológicos o de otra naturaleza.
7. **Sostenibilidad**. La sostenibilidad se refiere a la continuación de los beneficios de una **intervención** para el desarrollo después de concluida una asistencia externa importante; es una situación en la que las ventajas netas son susceptibles de resistir los riesgos con el correr del tiempo. Muchos ni siquiera consideran este aspecto, o bien asumen que los impactos se sostendrán solos. Sin embargo, los impactos pueden no sostenerse por una serie de razones: el proyecto puede recibir subsidios que serán suspendidos; la situación externa – por ejemplo, el clima o las circunstancias políticas o económicas - puede variar y por lo tanto amenazar la estabilidad de los efectos directos del proyecto; las causas subyacentes del problema pueden no haber sido abordadas, de manera tal que las intervenciones del proyecto no resolverán *estos* problemas y no serán sostenibles; o el proyecto puede introducir técnicas o tecnologías que las personas u organizaciones no están preparadas para manejar por sí solas.

⁴ Muchas de estas definiciones han sido tomadas del Glosario de la OCDE (Organización para la Cooperación Económica y el Desarrollo), 2002.

Figura 2 Modelo teórico simple de un programa



Todo proyecto es diseñado y ejecutado en un *contexto* singular, que incluye las circunstancias económicas, políticas, institucionales y ambientales locales y regionales, así como las características socioculturales de las comunidades o grupos afectados por el proyecto. La teoría del programa debe incorporar todos estos factores luego de realizar un *análisis contextual*. Cuando un proyecto es ejecutado en varias locaciones distintas, a menudo sucede que los resultados y efectos directos difieren significativamente entre un lugar y otro, debido a que las **variables contextuales** acusan diferentes configuraciones.

Adapte los planes de evaluación.

Las personas que encargan una evaluación deben considerar una serie de factores que deben ser incluidos en los términos de referencia. El cliente, y el evaluador/a que está siendo contratado para acometer la tarea, podrían valerse del siguiente grupo de preguntas para cerciorarse de que estos factores sean tomados en cuenta a medida que se elaboran los planes para llevar a cabo una evaluación. Las respuestas a estas preguntas pueden ayudar a focalizar la atención en los temas importantes en los que debe incidir la evaluación, entre ellos las formas de lidiar con las restricciones de una EMR.

- ¿Quién pidió la evaluación? ¿Quiénes son los interesados clave? ¿Tienen ideas preconcebidas sobre el propósito de la evaluación y los hallazgos previstos?
- ¿Quiénes deberían participar en la planificación/ejecución de la evaluación?
- ¿Cuáles son las preguntas clave que deben ser respondidas?
- ¿La evaluación será formativa o sumativa? ¿Su propósito principal es aprender y mejorar, rendir cuentas, o una combinación de ambos?
- ¿Habrá una segunda fase del proyecto o se diseñarán otros proyectos sobre la base de los hallazgos de esta evaluación?
- ¿Qué decisiones se tomarán en respuesta a los hallazgos de esta evaluación? ¿Quién las tomará?
- ¿Cuál es el nivel de rigor requerido para sustentar dichas decisiones?
- ¿Cuál es el campo de aplicación/el alcance de la evaluación?
- ¿Cuánto tiempo se requiere/de cuánto tiempo se dispone?
- ¿Qué recursos financieros se requieren/de qué recursos financieros se dispone?
- ¿Qué diseño de evaluación se requiere/es posible dadas las circunstancias?
- ¿La evaluación debería basarse principalmente en métodos cuantitativos (CUANTI), métodos cualitativos (CUALI) o una combinación de ambos?
- ¿Debemos emplear métodos participativos?
- ¿Puede/debe haber una encuesta de personas individuales, hogares u otras instituciones?
- ¿Quiénes deben ser entrevistados?

- ¿Qué diseño y tamaño de muestra se requieren/son factibles?
- ¿Qué forma de análisis responderá mejor a las preguntas clave?
- ¿Quiénes son los destinatarios del informe o informes? ¿De qué forma serán comunicados los hallazgos a cada grupo de destinatarios?

2. Selección del mejor diseño entre las opciones disponibles

La Tabla 2 identifica cinco estrategias de evaluación (diseños experimentales reales, diseños de campo aleatorios, diseños no aleatorios robustos, diseños no aleatorios débiles y diseños no experimentales) y la Tabla 3 resume siete diseños de EMR ampliamente utilizados que se emplean para operativizar dichas estrategias. La Tabla 4 analiza las fortalezas y debilidades de cada diseño. Los primeros dos diseños mencionados son los más robustos desde el punto de vista metodológico (los que están sujetos al menor número de amenazas a la validez de las conclusiones) y se recomiendan cuando el presupuesto y los plazos no son restricciones importantes y cuando la evaluación tiene la opción de empezar al inicio del proyecto. Los diseños 3 a 6 adolecen de limitaciones metodológicas importantes, pero pueden ofrecer estimados razonablemente robustos de los impactos de un programa y además pueden ser más flexibles en lo que respecta a la fecha de inicio. El diseño 7, que empieza recién cuando el proyecto se está aproximando a su fin y cubre únicamente al grupo del proyecto, es el más débil desde el punto de vista metodológico. Sin embargo, dadas las restricciones bajo las cuales se llevan a cabo muchas evaluaciones, se trata, lamentablemente, de la situación en la que se encuentran muchos evaluadores/as. Diseñado creativamente – utilizando diseños **de métodos mixtos**, que se valen de todos los enfoques de diseño, recolección de datos y análisis CUANTI y CUALI disponibles (ver el Capítulo 14) –, el Diseño 7 puede generar estimados operativamente útiles de la medida en que el proyecto contribuye a los efectos deseados. No obstante, dado el alto número de amenazas a la validez de las conclusiones, es importante revisar cuidadosamente las limitaciones del tipo de conclusiones que se pueden extraer del análisis. El Capítulo 10 describe cada uno de los diseños de evaluación, presenta un estudio de caso que ilustra cómo ha sido aplicado cada diseño en el campo y evalúa las principales amenazas a la validez e idoneidad de los enfoques. La Tabla 10.4* resume las condiciones bajo las cuales cada diseño funciona y no funciona adecuadamente. Ésta debe ser consultada en una etapa temprana de la planificación de la evaluación, a fin de ayudar a delimitar las opciones.

Más adelante describimos una serie de estrategias que pueden ser utilizadas para fortalecer la mayor parte de las EMR. Se recomienda enérgicamente a los evaluadores/as que consideren el uso de algunas de estas estrategias, dondequiera que sea apropiado y factible. La selección del diseño de EMR más apropiado depende de una serie de consideraciones, entre ellas las siguientes:

- ¿Cuándo empezó la evaluación? ¿Al inicio del proyecto, mientras el proyecto estaba siendo ejecutado o después que había concluido?
- ¿Cuándo terminará la evaluación? ¿Será una evaluación que se realice una sola vez, mientras el proyecto está siendo ejecutado (generalmente para la revisión de medio término), concluirá la evaluación

- aproximadamente al mismo tiempo que el proyecto (evaluación a la conclusión del proyecto e informe final), o continuará la evaluación después que concluya el proyecto (evaluación longitudinal o ex-post)?
- ¿*Qué tipo de comparación se usará?* Existen tres grandes opciones: (a) un diseño aleatorio en el cual personas individuales, familias, grupos o comunidades son asignados con carácter aleatorio a grupos del proyecto y **grupos de control**; (b) un **grupo de comparación** seleccionado porque coincide lo más cercanamente posible con las características del grupo del proyecto; o (c) ningún grupo de control o comparación de ningún tipo.
- ¿*El diseño incluye una evaluación de procesos?* Incluso si una evaluación se centra en la medición de los cambios sostenibles en la situación de la población objetivo, es importante identificar qué factores son los más susceptibles de haber generado dichos cambios. Esto incluye evaluar la calidad del proceso de ejecución del proyecto y si éste realizó una contribución plausible a cualquier impacto medido.
- ¿*Existen preferencias por el uso de enfoques CUANTI, CUALI o de métodos mixtos?* Ver los Capítulos 11, 12 y 13 al respecto.

Afine el diseño de evaluación.

Los enfoques de EMR se utilizan porque el tiempo, los recursos, los datos disponibles y, posiblemente, el contexto político no permiten el uso de diseños de evaluación más robustos. Bajo estas circunstancias, el evaluador/a debe acordar con el cliente cómo reducir los requerimientos de recursos y tiempo, así como las necesidades de datos, al mismo tiempo que se garantiza un nivel de precisión aceptable. Los Capítulos 3 y 4 analizan las opciones para realizar una evaluación dentro de las limitaciones de un presupuesto ajustado o bajo restricciones de tiempo, y el Capítulo 5 describe opciones para maximizar el uso de los datos disponibles. La Tabla 5 resume las alternativas para reducir los costos de recolección y análisis de datos, y la Tabla 6 estima el potencial ahorro de costos que entraña el uso de diseños de evaluación simplificados. Se recomiendan los siguientes pasos para definir el mejor y más aceptable diseño a la luz de determinadas restricciones:

Dedique el tiempo necesario a determinar claramente las necesidades prioritarias de información del cliente, así como las restricciones políticas y de otra índole bajo las cuales opera el proyecto:

- ¿Qué es lo que el cliente necesita realmente de la evaluación?
- ¿Es imprescindible realizar un riguroso análisis CUANTI para garantizar la credibilidad de la evaluación? ¿O un análisis CUALI en profundidad resultaría más importante y verosímil para los clientes?
- ¿Cuál es el nivel mínimo aceptable no negociable en términos de las necesidades de información y los plazos reales para producir un primer borrador y un informe final?

¿Quiénes son los interesados a los cuales está dirigida la evaluación y cuyas opiniones resultan fundamentales?

Revise las opciones para reducir los costos (Capítulo 3) y plazos (Capítulo 4), así como para fortalecer la base de datos disponible (Capítulo 5).

Con base en estas opciones, prepare varios escenarios posibles, que permitan alcanzar los objetivos de evaluación dentro de las restricciones de recursos existentes. Revise la “Lista de verificación estandarizada para evaluar la idoneidad y la validez de los diseños cuantitativos, cualitativos y de métodos mixtos” (Anexo 1), así como la “Lista de verificación para evaluar las amenazas a la validez de los diseños de evaluación cuantitativos” (Anexo 2) y sopesa las fortalezas y debilidades de cada opción desde la perspectiva del cliente o clientes y otros interesados clave.

Si ninguno de los escenarios disponibles satisface el nivel mínimo aceptable para el cliente, prepare dos escenarios adicionales:

Escenario 1: Calcule cuáles serían los requerimientos presupuestarios y cronológicos adicionales para satisfacer el nivel mínimo aceptable (por ejemplo: un monto adicional de \$25,000 para incluir un grupo de comparación en el diseño de evaluación, o una prórroga de tres meses en el plazo de entrega del informe preliminar).

Escenario 2: Indique qué modificaciones serían necesarias en el diseño de evaluación para poder trabajar con los recursos disponibles.

Bajo el escenario 2, una alternativa podría ser producir estimados consolidados del impacto del proyecto a nivel nacional, pero sin ofrecer estimados desagregados del impacto de diferentes combinaciones de servicios, o del impacto en diferentes sitios del proyecto, o en diferentes grupos socioeconómicos (por ejemplo, hombres y mujeres, trabajadores/as asalariados e independientes, grupos étnicos distintos).

Alternativamente, o adicionalmente, se podría reducir el nivel de confianza estadística (ver el Capítulo 14), de manera tal que las diferencias estadísticamente significativas entre los grupos del proyecto y de comparación podrían ser estimadas a un nivel de confianza de 0.10 (10 por ciento), y no al nivel convencional de 0.05 (5 por ciento). La Tabla 7 contiene una lista de los factores que afectan el tamaño de la muestra.

Una parte de la información podría recolectarse a través de, por ejemplo, entrevistas con **grupos focales** (entrevistas grupales focalizadas) y técnicas grupales de **DRP** (si bien este término se emplea ahora para denominar una amplia gama de métodos de diagnóstico participativo, originalmente significaba **diagnóstico rural participativo**), en vez de obtenerse mediante una encuesta muestral de hogares.

Es necesario analizar todas las opciones con el cliente, así como las implicaciones de cada opción en términos de: nivel de precisión, tipos de análisis que pueden efectuarse y credibilidad de los hallazgos para diferentes interesados. Antes de tomar una decisión sobre cómo proceder, es esencial cerciorarse de que el cliente entienda a cabalidad las opciones y concesiones. En ocasiones el cliente le preguntará al evaluador/a: “¿Qué enfoque recomendaría usted como el mejor, dado que usted es el experto/a?”. Si le hacen esta pregunta, el evaluador/a debe explicar que se trata de una decisión política que debe tomarse en consulta con el cliente, y que el rol del evaluador/a se limita a proporcionar asesoramiento acerca de las implicaciones técnicas de cada opción en términos de precisión, tipos de análisis y credibilidad profesional.

Identifique los tipos de análisis y comparaciones que resultan esenciales para la evaluación.

Un factor clave para la selección del diseño de evaluación, y para determinar el tamaño y la estructura de la muestra, tiene que ver con los tipos de análisis requeridos y los niveles de precisión estadística necesarios. Se recomienda considerar tres tipos de evaluaciones:

1. *Evaluaciones exploratorias o indagatorias*, cuyo propósito es determinar si el concepto y el enfoque básico del proyecto “funcionan”. Este tipo de evaluación se utiliza con frecuencia cuando se está introduciendo de manera piloto un nuevo tipo de servicio, o cuando se va a ofrecer un servicio ya existente pero de otra manera, o para llegar a nuevos grupos objetivo. Algunos ejemplos de preguntas clave para este tipo de evaluación son:
 - a. ¿Están dispuestos los agricultores a experimentar con nuevos tipos de semillas?
 - b. ¿Están obteniendo los nuevos métodos de enseñanza una respuesta positiva de parte de las escuelas, los alumnos/as y los padres y madres de familia, y existe alguna **evidencia** inicial de que el rendimiento escolar haya mejorado?
 - c. Si se otorgan préstamos a las mujeres pobres, ¿podrán éstas utilizar el dinero para empezar un pequeño negocio o expandirlo?
 - d. ¿Qué grupos se benefician más del proyecto y qué grupos menos y por qué?
2. *Diseños cuasi-experimentales o CUALI a pequeña escala*, que se utilizan para evaluar si existen evidencias de que el proyecto está generando efectos significativos entre la población objetivo. Algunos diseños incluyen un grupo de comparación, mientras que otros realizan una comparación más general con comunidades similares, empleando técnicas de DRP o entrevistas con grupos focales. Los aspectos relacionados con la atribución de los efectos directos (¿cuál hubiera sido la situación del grupo afectado por el proyecto si el proyecto no hubiese tenido lugar?) se tocan pero de una manera menos rigurosa que en el caso de las evaluaciones de impacto a gran escala (ver descripción

más abajo). Algunas de las preguntas fundamentales para este tipo de evaluación podrían ser:

- a. ¿Los beneficiarios previstos del proyecto (por ejemplo, personas individuales, familias, escuelas, comunidades) se encuentran en mejor situación ahora como consecuencia del proyecto?
 - b. ¿Qué tan seguros estamos de que los cambios observados han sido causados por el proyecto y no por factores externos, por ejemplo, la mejora de la economía local?
 - c. ¿El proyecto tendría probabilidades de causar efectos similares en otras áreas si fuese replicado? ¿Dónde funcionaría bien y dónde menos bien y por qué?
 - d. ¿Qué factores contextuales (económicos, políticos, institucionales, medioambientales y culturales) afectan el éxito del proyecto? (Ver discusión en el Capítulo 9 sobre factores contextuales y variables mediadoras.)
 - e. ¿Quiénes se han beneficiado y quiénes no se han beneficiado del proyecto?
3. *Evaluación de impacto a gran escala*, cuyo propósito es evaluar con mayor rigor estadístico la magnitud del efecto producido (definida numéricamente en términos de porcentajes o cambios cuantitativos) y quiénes se benefician y quiénes no. Idealmente se debe emplear para este tipo de evaluación un enfoque de métodos mixtos, que integre métodos CUANTI y CUALI. Algunas preguntas críticas para este tipo de evaluación podrían incluir:
- a. ¿Qué impactos CUANTI (efectos sostenibles de alto nivel) han sido generados por el proyecto? El énfasis recae en el “cuánto” y no solamente en el “qué”.
 - b. ¿Cuál es la calidad de los servicios (en comparación con otros programas y con la norma esperada, así como en opinión de los grupos objetivo)?
 - c. ¿Los efectos del proyecto son estadísticamente significativos “más allá de toda duda razonable”?
 - d. ¿Quiénes se han beneficiado más y quiénes menos, y existen grupos que no se hayan beneficiado en absoluto o que incluso estén en peor situación que antes?
 - e. ¿Cuáles son las **variables de intervención** (por ejemplo, las características socioeconómicas de los grupos afectados por el proyecto, los factores culturales que afectan la participación) que influyen en la magnitud de los impactos?

En las *evaluaciones exploratorias* probablemente bastaría con un análisis descriptivo basado en técnicas tales como: observación, entrevistas con al menos unos cuantos miembros seleccionados de la población objetivo, entrevistas con informantes clave y quizás entrevistas con grupos focales. También se podría efectuar una encuesta rápida y sencilla, destinada a recolectar información básica sobre la población afectada por el proyecto. Probablemente no sería necesario utilizar un grupo de comparación

formal, aunque se podrían realizar visitas a otras comunidades o áreas similares para evaluar cuán similares o diferentes son en comparación con las áreas del proyecto.

En las *evaluaciones CUANTI y CUALI a pequeña escala de los efectos directos e impactos* resulta útil, aunque no siempre posible, identificar un grupo de comparación, que ayudará a estimar cuál hubiera sido la situación del grupo del proyecto si el proyecto no se hubiese producido (el **contrafactual**). Idealmente se utiliza un enfoque de métodos mixtos, evaluando tanto la calidad como la cantidad de los servicios e impactos. Además, si se pueden obtener datos de base (cualquiera sea la forma en que se encuentren disponibles), el diseño será mucho más robusto. Se deben realizar comparaciones estadísticas simples entre grupos del proyecto y de comparación, tales como las diferencias entre las medias o porcentajes.

En el caso de las *evaluaciones de impacto a gran escala* se requieren muestras relativamente grandes (y a menudo cientos de observaciones de grupos tanto del proyecto como de comparación), de manera tal que se puedan emplear análisis multivariados para controlar estadísticamente las diferencias entre los grupos del proyecto y de comparación y estimar la influencia cuantitativa de las variables de intervención. Aquí nuevamente debe usarse un enfoque de métodos mixtos, de manera tal que los estimados CUANTI sean complementados por descripciones CUALI del contexto del proyecto, el proceso de ejecución del proyecto, la calidad de los servicios y las opiniones y experiencias de los beneficiarios e instituciones y miembros del personal responsables de la ejecución del proyecto.

Evalúe las amenazas a la validez y la idoneidad de diferentes diseños.

El enfoque de EMR evalúa las fortalezas y debilidades de las diferentes etapas del diseño de evaluación con el fin de identificar los factores que afectan la validez de las conclusiones y recomendaciones. Se trata de un aspecto importante para cualquier evaluación, pero especialmente para una EMR, donde a menudo se tienen que relajar los procedimientos metodológicos convencionales, ya sea debido a restricciones cronológicas o presupuestarias o porque una parte de los datos requeridos no se encuentra disponible. Existen varios factores que afectan la idoneidad del diseño y los hallazgos de una evaluación (ver la Tabla 11 y el Capítulo 7 al respecto), incluyendo los siguientes:

- Idoneidad de la focalización, el enfoque y los métodos de evaluación utilizados para obtener los tipos de información requeridos
- Disponibilidad de los datos y fuentes de datos
- Capacidad de los datos recolectados para respaldar las interpretaciones acerca de los resultados e impactos del programa
- Calificaciones del equipo de evaluación, en términos tanto de la metodología de evaluación como de los campos específicos en los que incide el programa

En el caso de las evaluaciones CUANTI, Cook y Campbell establecieron en la década de 1960 cuatro grupos de “amenazas a la validez” que gozan de aceptación general (ver versión actualizada en Shadish, Cook y Campbell, 2002). El Anexo 2

(Secciones F, G, H e I, respectivamente) describe y explica los diversos tipos de amenazas a los diseños de evaluación CUANTI para cada una de las cuatro categorías que se describen a continuación:

Amenazas a la validez de la conclusión estadística (Anexo 2, Sección F) – inferencias/conclusiones incorrectas sobre los efectos de las intervenciones del proyecto en los resultados e impactos previstos: Los problemas pueden surgir debido a la aplicación incorrecta de una prueba estadística o a las limitaciones de la muestra (por ejemplo, si la muestra no es representativa de toda la población). Otro problema común es que, debido a lo reducido del tamaño de la muestra (generalmente por restricciones presupuestarias o cronológicas), el análisis puede concluir que el proyecto no ha tenido un efecto estadísticamente significativo, cuando en realidad la muestra era demasiado pequeña como para constatar dicho efecto, incluso si éste se ha producido. A menudo se realizan concesiones mutuas entre reducir el tamaño de la muestra con el fin de reducir los costos, de un lado, y la necesidad de garantizar que la muestra sea lo suficientemente grande como para detectar los efectos, si es que existen, del otro. Por otra parte, muchos proyectos bien diseñados y ejecutados solamente aspiran poder producir efectos leves, lo cual dificulta aún más la detección de los efectos. En el Capítulo 14 se analizan los problemas relacionados con errores de Tipo II (concluir equivocadamente que no se produjeron efectos), el **poder de la prueba, el tamaño del efecto**, y la estimación del tamaño de la muestra.

Amenazas a la validez interna (Anexo 2, Sección G) – inferir de manera incorrecta que la intervención del proyecto “causó” o contribuyó a un efecto directo o impacto en particular: Muchos problemas relacionados con la validez interna se derivan de la forma en que se selecciona a los participantes del proyecto (que acusan diferencias importantes con el grupo de comparación), dado que las características del grupo del proyecto varían a lo largo de la vida del proyecto, por ejemplo cuando hay personas que abandonan el proyecto o cuando algunas experiencias durante la ejecución del proyecto influyen en la respuesta de los participantes.

Amenazas a la validez del marco conceptual (Anexo 2, Sección H) – es decir, el grado al que se pueden realizar inferencias legítimas a partir de los marcos conceptuales teóricos (la definición de los conceptos clave) en los cuales se basa la teoría del programa: Muchos de los marcos conceptuales clave son difíciles de definir (por ejemplo, pobreza, vulnerabilidad, bienestar, ambiente de trabajo hostil), y más difíciles aún de medir. La falta de precisión y claridad en la definición y medición de dichos marcos conceptuales clave socavarán la capacidad

de la evaluación para **determinar** e interpretar la forma en que se desempeñó el proyecto y los resultados que se obtuvieron.

Amenazas a la validez externa (Anexo 2, Sección I) – inferencias incorrectas sobre la probabilidad de que los hallazgos de la evaluación sean aplicables a diferentes personas, tiempos o contextos: La mayor parte de las evaluaciones CUANTI está diseñada para determinar la medida en que los hallazgos de la evaluación pueden ser generalizados y aplicados a una población más amplia (por ejemplo, todas las comunidades de bajos ingresos, todas las trabajadoras no calificadas, todos los niños y niñas de la edad para educación secundaria). Hay una serie de aspectos en los que las características de la población del proyecto pueden no ser representativas de la población más amplia; por lo tanto, la generalización de los hallazgos puede llevar a conclusiones equívocas. Por ejemplo, un programa de alfabetización de adultos puede haber sido exitoso, al menos en parte, gracias al entusiasta apoyo de la cámara de comercio local, que proporcionó el transporte, cuadernos de ejercicios gratuitos y refrigerios para los participantes. Por consiguiente, el éxito del programa no justifica la conclusión de que éste sería igualmente exitoso en otras ciudades, donde es posible que no se cuente con el mismo respaldo local.

La lista de verificación de amenazas a la validez de los diseños puede utilizarse para identificar y evaluar las potenciales debilidades de los siete diseños CUANTI de EMR más frecuentes (ver Capítulos 7 y 10 al respecto). Los lectores/as no especializados en análisis estadístico y evaluación CUANTI posiblemente encontrarán las secciones F a I de la lista de verificación integral difíciles de seguir. Aunque vale la pena echar una mirada a las categorías para formarse una idea de la amplia gama de factores que pueden afectar la validez de los diseños de evaluación CUANTI, siempre queda la posibilidad de consultar con un/a especialista a la hora de diseñar y evaluar una evaluación CUANTI rigurosa.

Para los lectores/as no especializados en el análisis estadística, las secciones A hasta E del Anexo 1 (ver información más adelante) proporcionan pautas suficientes para evaluar la validez de la mayor parte de las evaluaciones. Éstas suelen emplearse para evaluar la validez (a menudo conocida también como idoneidad) de los diseños de evaluación CUANTI o de métodos mixtos, pero también complementan las categorías F a I (ver Anexo 2) en lo que se refiere a la evaluación de diseños de evaluación CUANTI.

Confirmabilidad y objetividad (Anexo 1, Sección A): ¿Las conclusiones han sido extraídas de y/o se sustentan en la evidencia disponible en los informes de la evaluación, y se encuentra relativamente libre de prejuicios la investigación por parte del investigador/a? Hay cinco subcategorías para evaluar cuán bien han sido descritos los métodos y procedimientos, si se han aportado datos para corroborar las conclusiones, si se reconocen y toman en cuenta los potenciales

prejuicios del investigador y si se han considerado hipótesis divergentes.

Confiabilidad y fiabilidad (Anexo 1, Sección B): ¿El proceso del estudio se mantiene constante, coherente y razonablemente estable a través del tiempo y en lo que concierne a todos los investigadores/as y métodos? Hay ocho subcategorías que se refieren a la claridad de las preguntas de investigación, la aplicación de los métodos de recolección de datos y los procedimientos para verificar y validar dichos métodos. Mientras que las evaluaciones CUANTI intentan garantizar la uniformidad y la estandarización, muchas evaluaciones CUALI emplean *diseños emergentes* ahí donde los métodos de recolección e interpretación de datos evolucionan y cambian a medida que progresa la evaluación. El desafío radica en evaluar si los procesos emergentes se encuentran adecuadamente documentados y evaluados, de manera tal que el lector/a pueda entender la base de evidencia en la que se sustentaron las conclusiones.

Validez interna, credibilidad y autenticidad (Anexo 1, Sección C): ¿Son creíbles los hallazgos para las personas que están siendo estudiadas y para los lectores/as, y ofrecen un retrato veraz de lo que está siendo estudiado? Las ocho subcategorías inciden en la profundidad y la riqueza de la recolección y el análisis de datos, cómo perciben diferentes grupos la credibilidad del análisis, la coherencia de los hallazgos con la teoría y las hipótesis existentes, y si los hallazgos se encuentran debidamente respaldados por datos.

Validez externa, transferibilidad y generalización (Anexo 1, Sección D): ¿Las conclusiones son adaptables a otros contextos y cuán ampliamente pueden ser generalizadas? Las 11 subcategorías evalúan cuán ampliamente pueden ser generalizados los hallazgos y a qué otros grupos o poblaciones.

Utilización, aplicación y orientación a la acción (Anexo 1, Sección E): ¿Cuán útiles fueron los hallazgos para el cliente, los investigadores/as y las comunidades estudiadas? Las siete subcategorías proporcionan indicadores para evaluar la utilidad.

La Tabla 12 identifica algunas de las amenazas a la validez que pueden afectar hasta los diseños de evaluación más robustos.

Cuándo y cómo utilizar la “lista de verificación de amenazas a la validez y la idoneidad”

La lista de verificación debe ser consultada a todo lo largo de la evaluación, para cerciorarse de que las cosas estén yendo por el camino correcto e identificar y resolver rápidamente los problemas que estén afectando la validez de la metodología y las conclusiones. Los momentos más importantes para consultar la lista de verificación son:

Cuando se está diseñando la evaluación y se están planeando los métodos de recolección y análisis de datos. La lista de verificación puede

emplearse para identificar potenciales problemas y considerar alternativas para resolverlos. En la EMR, es especialmente útil para identificar las potenciales amenazas a la validez que surgen de los métodos propuestos para reducir los costos y plazos y para trabajar con bases de datos limitadas. Por ejemplo, ¿qué amenazas adicionales surgen para la validez si se elimina el grupo de comparación de base o se recolectan datos sobre el ingreso a través de entrevistas con grupos focales en vez de encuestas muestrales de hogares?

Una vez recolectada la mayor parte de los datos. La lista de verificación puede ayudar a identificar potenciales amenazas que hayan surgido durante la recolección de los datos (por ejemplo, tasas inesperadamente altas de preguntas sin responder o confusión en torno al concepto de desempleo). Debe aplicarse lo antes posible después de concluida la recolección de datos (o incluso cuando ésta todavía está en marcha), para estar a tiempo de tomar medidas correctivas.

Una vez concluido el informe de evaluación preliminar. Idealmente, aún se estará a tiempo para tomar algunas medidas correctivas. Sin embargo, si esto no fuera posible, el informe de evaluación deberá emplear la lista de verificación para identificar y señalar claramente las potenciales amenazas a la validez de los hallazgos y de qué forma podrían afectar las conclusiones y recomendaciones.

El Anexo 2* de Evaluación en un Mundo Real (que no está incluido en esta visión general) contiene una “Hoja de trabajo para proyectos de EMR, para identificar y abordar las amenazas a la validez y la idoneidad” que puede ser utilizada en cualquier etapa de la evaluación a fin de identificar las amenazas a la validez e idoneidad de la evaluación, evaluar la seriedad de cada amenaza para los propósitos de la evaluación e identificar las acciones que pueden ser adoptadas para corregir o al menos paliar las amenazas más importantes. El Anexo 3* presenta un ejemplo elaborado que ilustra cómo podría aplicarse la mencionada hoja de trabajo para evaluar la evaluación de un proyecto de vivienda.

3. Determinación de la metodología apropiada

Los enfoques y métodos CUANTI y CUALI suelen ser diseñados con propósitos diferentes. Los Capítulos 11 y 12 inciden en las fortalezas y potenciales debilidades de los enfoques CUANTI y CUALI, respectivamente. Reconociendo estas fortalezas y debilidades, el enfoque de EMR considera que en la mayoría de los casos el diseño de evaluación más sólido y robusto probablemente sea una combinación de enfoques tanto CUANTI como CUALI. El Capítulo 13 contiene una discusión detallada de cómo los métodos mixtos, que combinan sistemáticamente enfoques CUANTI y CUALI, pueden ser útiles para una EMR. Sostiene que los enfoques de métodos mixtos son especialmente valiosos para una EMR porque el uso de diversos métodos de recolección y análisis de datos permite detectar y eliminar, o al menos reducir, algunas de las amenazas a la

validez que se generan a raíz de los compromisos que se tienen que hacer a la luz de las restricciones presupuestarias, cronológicas y de información. Por ejemplo, reducir el tamaño de la muestra incrementa el margen de error del muestreo, dificultando aún más la detección de diferencias significativas. Si se emplean técnicas de DRP y otras técnicas CUALI para obtener estimados independientes de los impactos del proyecto, y si los hallazgos son coherentes con el análisis estadístico de las encuestas muestrales, aumentará la confianza en los hallazgos.

Muchos autores/as sostienen que, cuando se emplean enfoques de métodos mixtos, todas las evaluaciones tienen un enfoque predominante, ya sea CUANTI o CUALI (“inclinación teórica”), y utilizan el segundo enfoque para complementarlo. La inclinación teórica dependerá de la orientación profesional del investigador/a o de las preferencias del cliente. Sin embargo, otros autores/as proponen un enfoque integrado, que no privilegie ninguno de los dos enfoques. Sea cual fuere la posición que se adopte en este tema, algunos programas se prestan más naturalmente al uso de métodos de evaluación CUANTI (por ejemplo, programas muy grandes que afectan a miles de personas), mientras que en otros casos los métodos CUALI parecen ser los más indicados (por ejemplo, un programa cuya meta es mejorar la calidad de las prácticas de enseñanza, o un programa que esté tratando de introducir el cultivo de verduras en diversas aldeas relativamente pequeñas). Sin embargo, en todos los casos la selección de los métodos se verá influenciada también por las preferencias de los evaluadores/as y los clientes.

A todo lo largo de este libro invocamos fuertemente a los evaluadores/as a seleccionar las herramientas de recolección y análisis de datos más apropiadas para las necesidades del cliente y para la naturaleza del programa que está siendo evaluado, y evitar seleccionar los métodos simplemente en razón de su naturaleza cualitativa o cuantitativa.

4. Opciones para fortalecer los diseños de EMR

Debido a los contextos en los cuales se realiza una EMR, los diseños CUANTI convencionales, cuasi-experimentales, utilizados de manera aislada, son vulnerables a una serie de amenazas que tienden a debilitar la calidad de los datos y la validez de las conclusiones. (Es por eso que la EMR recomienda fuertemente el uso de diseños de métodos mixtos.) Estos aspectos se tratan en mayor detalle en el Capítulo 10. Las limitaciones de los diseños CUANTI convencionales incluyen:

- Pueden surgir problemas en relación con la confiabilidad de la medición de indicadores clave, en especial cuando se incide en temas sensibles como el consumo ilegal de estupefacientes, el control de los recursos dentro del grupo familiar, la violencia doméstica y las restricciones sociales a las actividades o la movilidad de la mujer.

- Puede haber dificultades para detectar las variaciones en la ejecución del proyecto, la calidad de los servicios y el acceso de distintos grupos a los servicios y beneficios.

Los diseños convencionales no analizan los factores contextuales que afectan los efectos directos e impactos del proyecto en diferentes localidades.

Las diferencias importantes (no equivalencia) entre los grupos del proyecto y de comparación son difíciles de detectar, especialmente cuando son difíciles de cuantificar (por ejemplo, motivación, organización comunitaria).

A continuación presentamos una serie de procedimientos que pueden emplearse para fortalecer los diseños referidos (ver también el Recuadro 7.2*, Capítulo 7, para ejemplos de formas de enfrentar amenazas comunes a la validez estadística interna, del marco conceptual y externa). Los evaluadores/as deben considerar incorporar algunos de los siguientes procedimientos a los diseños de EMR ahí donde proceda.

Sustente el diseño de evaluación en el modelo teórico [marco lógico] del programa.

Como hemos visto anteriormente en este capítulo, la formulación del modelo teórico del programa ayuda a identificar los temas e hipótesis clave en los cuales deberían enfocarse los limitados recursos de evaluación (ver Capítulos 2 y 9). El modelo teórico puede utilizarse, además, para complementar un diseño cuasi-experimental - describiendo el proceso de ejecución del proyecto y analizando los factores contextuales que afectan la ejecución y los efectos directos del proyecto - y ayuda a interpretar los hallazgos de la evaluación y a evaluar si un proyecto debería continuar o ser replicado.

Complemente la evaluación cuantitativa (sumativa) con una evaluación de procesos.

La **evaluación de procesos** se vale de métodos CUANTI, CUALI y mixtos para observar y evaluar el proceso de ejecución de un proyecto y formular recomendaciones para fortalecer las fases subsiguientes de un programa en ejecución. Responde a preguntas tales como las siguientes:

- ¿De qué forma se ejecutaron los diferentes componentes del proyecto, y qué tan estrechamente se ciñó la ejecución del proyecto sobre el terreno al plan del proyecto o al manual operativo del mismo?
- En proyectos en ejecución, ¿cómo se podría evaluar y mejorar la calidad de los servicios? ¿Existen evidencias de que éstos están produciendo los efectos directos deseados?
- ¿Quién tiene acceso a y/o utiliza los servicios y quién no? ¿Por qué ciertos grupos no utilizan los servicios?
- ¿El diseño y la organización del proyecto fueron participativos, dirigidos por un grupo pequeño o impuestos desde arriba? ¿Quiénes participan en la toma de decisiones durante la ejecución del proyecto?
- ¿Qué porcentaje de la comunidad (beneficiarios previstos) sabe del proyecto? ¿Su información es correcta? ¿Qué opinan del proyecto?
- ¿Cuál es la relación entre los organizadores del proyecto y la comunidad?

¿Qué saben las organizaciones gubernamentales y otras organizaciones acerca de este proyecto, y qué opinan de la calidad de los servicios y la eficacia del proyecto?

Incorpore el análisis contextual.

El análisis contextual evalúa la influencia de los factores económicos, políticos, institucionales y medioambientales en la ejecución y el efecto directo de los proyectos. Dichos factores son definidos en los modelos teóricos de los programas como factores **mediadores** (ver el Capítulo 9). Además, examina la influencia de las características socioculturales preexistentes de las poblaciones objetivo en la respuesta de distintos grupos al proyecto.

Si bien la mayor parte del análisis contextual es cualitativa (por ejemplo, entrevistas con informantes clave, revisión de la documentación del proyecto y **observación participante**), éste puede incluir también un análisis CUANTI de los datos recogidos en encuestas de hogares. Las **variables contextuales** también pueden ser transformadas en variables dicotómicas (por ejemplo, “**variables dummy**”) e incorporadas en un análisis multivariado (ver el Capítulo 13, nota 1).

Reconstruya la situación de base.

Muchas evaluaciones empiezan recién cuando el proyecto lleva algún tiempo en ejecución, o quizás incluso cuando se está aproximando al final. En estas circunstancias, es muy común descubrir que no se recolectaron datos de base al inicio del proyecto. Esto es más común en el caso del grupo de comparación, pero a menudo es cierto también para el grupo del proyecto. La ausencia de una línea de base suele ser una de las amenazas más graves a la validez, y por lo tanto la EMR propone una serie de opciones para *reconstruir* la situación de base. Algunos de estos enfoques, que se describen en la Tabla 8 (ver también el Capítulo 5), incluyen:

Uso de datos secundarios (ver la siguiente sección)

Uso de métodos de retrospectiva individual (se pide a los entrevistados que recuerden la situación de su familia o comunidad alrededor de la fecha en que se inició el proyecto)

Uso de técnicas de DRP y otras técnicas participativas para *reconstruir* la historia de la comunidad y evaluar los cambios que han sido generados por el proyecto

Entrevistas con informantes clave, de preferencia personas que conocen a la comunidad objetivo así como a otras comunidades y por lo tanto poseen una perspectiva de los cambios relativos que se han producido a través del tiempo

Aunque todos estos métodos generan información potencialmente valiosa, existen amenazas considerables a la validez que son inherentes a cualquier método de retrospectiva. Éstas tienen que ver con la falta de recuerdos precisos, la tendencia a confundir el período de tiempo exacto (de manera tal que eventos que tuvieron lugar con

anterioridad pueden ser reportados como si hubieran ocurrido después que empezó el proyecto o viceversa) y, en algunos casos, esfuerzos deliberados de distorsión. En consecuencia, es importante tratar con cautela todos los datos obtenidos por retrosección y utilizar siempre enfoques de métodos mixtos para triangular los estimados independientes de información obtenida de distintas fuentes.

Uso de datos secundarios

Planear y monitorear administrativamente los datos recolectados por la organización que está siendo evaluada. Los datos recolectados como parte de un diagnóstico anterior al proyecto, así como los datos recogidos por el sistema de seguimiento de los ejecutores del proyecto durante la vida del proyecto, constituyen fuentes importantes pero a menudo subutilizadas para reconstruir la situación de base (ver el Capítulo 5). La mayoría de los proyectos recolecta una gran cantidad de datos con fines administrativos y de seguimiento, y dichos registros con frecuencia contienen información que puede ser útil para reconstruir la información sobre la situación de la población del proyecto al momento de empezar el proyecto (datos de base). Algunos de los tipos de información que típicamente se recogen son:

- Estudios de planificación y de factibilidad previos al inicio del proyecto
- Características socioeconómicas de las personas o familias que solicitan o reciben servicios
- Asistencia a reuniones comunitarias así como, posiblemente, informes de las reuniones (por ejemplo, las actas)
- Informes de actividades elaborados por el personal de la entidad ejecutora u otras personas involucradas en la ejecución del proyecto. Éstos ofrecerán, como mínimo, información sobre el proceso, pero idealmente mencionarán también los cambios observados en el conocimiento, las actitudes y las prácticas de los clientes.

Si bien estos tipos de datos administrativos pueden ser extremadamente útiles para reconstruir la situación de base, es importante tener en cuenta que los datos no fueron recolectados con fines de evaluación y por lo tanto pueden tener algunas limitaciones (por ejemplo, pueden estar incompletos o mal mantenidos o pueden no incluir toda la información requerida para los fines de la evaluación). La segunda sección del Capítulo 5 contiene una relación de algunas de las preguntas que deberían ser formuladas para evaluar la calidad y la utilidad de dicha información para reconstruir los datos de base.

Registros de otras organizaciones. Los registros de otros programas o proyectos que se encuentran presentes en la misma zona muchas veces ofrecen información sobre la situación anterior al inicio del proyecto. Por ejemplo, a menudo se realizan encuestas para estimar el número de niños y niñas que no están en la escuela, las fuentes y el costo del suministro de agua o la disponibilidad de microcrédito. Pueden disponer además de información estadística más general sobre, por ejemplo, las tasas de matrícula escolar, el

índice de mortalidad infantil, los precios de productos agropecuarios, los préstamos en la modalidad de microcrédito y los patrones del transporte. Es importante evaluar las fortalezas y debilidades de dichos registros en los siguientes sentidos:

- Diferencias entre la fecha de inicio del proyecto (que es cuando se requieren datos para el estudio de base) y la fecha en que se recogieron datos secundarios. Las diferencias en los tiempos son particularmente importantes cuando la situación económica general ha variado entre la fecha de la encuesta y la fecha de inicio del proyecto.
- Diferencias en términos de la población cubierta. Por ejemplo, ¿las encuestas incluyeron datos sobre el empleo en el sector informal tanto como formal? ¿Incluyeron los medios de transporte pedestre tanto como vehicular?
- ¿Se recolectó información sobre variables clave y sobre los potenciales impactos del proyecto? ¿Son estadísticamente válidos los datos secundarios para la población objetivo a la que supuestamente beneficiará el proyecto objeto de evaluación?
- ¿La información se refiere tanto a los hombres como a las mujeres? ¿O toda la información se obtuvo de una sola persona (generalmente el “jefe de hogar”) y se consolidó para todos los miembros del hogar (ver el Recuadro 5.2)?

Use enfoques de métodos mixtos para fortalecer la validez de los indicadores y mejorar la interpretación de los hallazgos.

Los siete diseños de evaluación que se describen en el Capítulo 10 pueden ser fortalecidos utilizando diseños de métodos mixtos (ver el Capítulo 13) que combinen enfoques CUANTI y CUALI en una de las siguientes modalidades:

- Estudios exploratorios para determinar el contexto e identificar los problemas clave y las hipótesis a ser validadas. Esto es especialmente importante para la construcción del modelo teórico del programa (ver el Capítulo 9);
- Análisis de la calidad de los servicios provistos por el proyecto;
- Análisis de la accesibilidad del proyecto para distintos sectores de la población objetivo;
- Análisis de los factores contextuales (la situación económica, política, institucional y ambiental natural) dentro de los cuales opera cada sitio del proyecto;
- Determinación de las características culturales de las poblaciones afectadas y de la forma en que dichos atributos influyen en la ejecución y los efectos directos del proyecto;
- Uso de la **triangulación** para ofrecer dos o más estimados independientes de indicadores clave tanto de procesos como de efectos (ver más adelante).

Los estimados siempre serán más sólidos si pueden ser confirmados de manera independiente de dos o más fuentes independientes. Esto se puede hacer de la siguiente manera:

Utilizando estimados independientes de cambios en las variables de impacto, obtenidos por medio de estudios basados en observaciones, entrevistas con grupos focales y datos secundarios. Comparar los estimados mediante triangulación. Si los estimados de fuentes distintas coinciden entre sí, esto permitirá tener mayor confianza en los hallazgos.

Si los estimados no coinciden, se deberá contar con una estrategia de seguimiento ex post para determinar las razones y realizar ajustes en los estimados.

El Capítulo 10 ilustra el uso de la triangulación para comparar tres estimados independientes (encuesta, observación e informantes clave) del ingreso al nivel de hogares. En un ejemplo, los tres estimados son *convergentes* (coinciden entre sí), mientras que en el segundo ejemplo los estimados son *divergentes* (no coinciden entre sí). En el segundo caso será necesario hacer un seguimiento ex post para determinar la razón de la falta de coincidencia y decidir cómo seleccionar el valor (o valores) más creíble(s) que han de ser utilizados en el análisis. Idealmente, el diseño de evaluación debería programar el tiempo y los recursos necesarios para regresar al campo y hacer un seguimiento ex post, ya sea durante la fase de supervisión de las entrevistas o durante la fase de análisis (que es cuando tienden a salir a la luz las discrepancias). Lamentablemente, éste no suele ser el caso, especialmente a la luz de las restricciones presupuestarias a las que está sometida una EMR. Por consiguiente, será necesario considerar otras opciones:

Instruir a los **encuestadores/as** que anoten los casos en que la información reportada y su observación directa no coincidan entre sí. Indicarles que expliquen cuál es su interpretación de las discrepancias y posiblemente cuál estimado es el que consideran más verosímil. También se les puede instruir que hagan algunas preguntas de seguimiento ex post (ver el Capítulo 11).

Identificar las inconsistencias durante la fase de supervisión de las entrevistas y hacer un seguimiento ex post, que consistirá de una discusión con los encuestadores/as pos-entrevistas y posiblemente nuevas visitas a una muestra de encuestados.

Normas para la codificación del instrumento de la encuesta y análisis de cómo abordar las inconsistencias (por ejemplo, ¿debe darse mayor peso a una fuente de información que a otra? ¿los estimados CUANTI deben ser adaptados de alguna forma determinada?).

En el caso de variables críticas, se pueden crear dos indicadores diferentes dando estimados superiores e inferiores (por ejemplo, ingreso, matrícula escolar, desempleo): uno en el cual la información de la

encuesta no haya sido adaptada y el otro donde sí lo haya sido. Ambos estimados serán presentados por separado en el análisis.

La Tabla 13 resume las características de los enfoques cuantitativos y cualitativos en diferentes etapas del proceso de evaluación, mientras que la Tabla 14 describe los elementos de un diseño de investigación multidisciplinario integral.

Controle las diferencias entre los grupos del proyecto y de comparación.

Análisis multivariado

Cuando se realizan encuestas con una muestra extensa, a menudo se utiliza el análisis multivariado con el fin de controlar estadísticamente las diferencias entre los grupos del proyecto y de comparación y así fortalecer los estimados del impacto del proyecto (ver Capítulos 10 y 11). El análisis realiza un empareamiento estadística entre los sujetos (por ejemplo, personas, hogares) en variables tales como edad, educación e ingreso, y determina si sigue habiendo una diferencia importante entre los grupos del proyecto y de comparación en lo que respecta a la variable del impacto (por ejemplo, porcentaje de niños y niñas que asisten a la escuela, número de adultos desempleados). Si una vez que se haya controlado estadísticamente por esa variable sigue habiendo una diferencia significativa, aumentará la confianza en que el proyecto realmente está contribuyendo a la diferencia. No obstante, si la diferencia desaparece, esto sugeriría que la asistencia escolar o el desempleo pueden estar determinados en mayor medida por las características de los hogares que por la participación de los mismos en el proyecto.

Uso de métodos CUALI para analizar las diferencias entre los grupos del proyecto y de comparación

Los informantes clave, las entrevistas con grupos focales y la observación participante son algunos de los métodos que se pueden utilizar para identificar diferencias potencialmente importantes entre los grupos del proyecto y de comparación, y para evaluar cómo éstas podrían afectar los efectos directos del proyecto y la estimación de los impactos del mismo. (Ver el Capítulo 5 sobre recolección de datos para llegar a grupos difíciles de acceder y métodos rápidos para comparar a los grupos del proyecto con grupos de comparación.)

5. Economizar en recursos humanos para la evaluación

En esta sección tratamos algunos temas referentes a la contratación de expertos/as externos (ya sea de otro país o de otra región dentro del mismo país), especialistas en áreas de contenido y recolectores/as de datos disponibles a nivel local. Lo ideal es formar un equipo de evaluación que incluya una adecuada combinación de personas con diferentes experiencias, destrezas y perspectivas. Cuando se enfrentan restricciones, especialmente financieras, como en el caso de una EMR, posiblemente se tengan que hacer compromisos en lo que concierne a la composición del equipo de evaluación.

Aunque en esta sección nos referimos a cada una de estas categorías de personas por separado, lo importante es considerar la combinación general y la eficacia del equipo de evaluación en su conjunto para satisfacer los requisitos de una evaluación.

Utilice consultores/as externos con inteligencia.

Los consultores/as externos suelen ser contratados: (a) debido a la falta de experticia técnica a nivel local (tanto al interior de la organización como en la comunidad de investigación local), (b) con el fin de desarrollar capacidades locales, (c) para ahorrar tiempo, (d) para garantizar independencia y objetividad, (e) para garantizar la credibilidad de la evaluación, y/o (f) porque se trata de un requisito de la entidad financiera. Si bien los consultores/as externos pueden mejorar sustancialmente la calidad de las evaluaciones tanto presentes como futuras si son correctamente seleccionados y utilizados, también son caros y en ocasiones pueden crear conflictos; en consecuencia, deben ser seleccionados y utilizados con inteligencia. Bajo las restricciones de una EMR, el objetivo debe ser limitar el uso de consultores/as externos a aquellas áreas en las que resultan indispensables. Aunque muchos de los aspectos que tratamos a continuación se refieren al uso de especialistas en evaluación nacionales e internacionales en países en desarrollo, los mismos principios rigen para los países desarrollados. Por ejemplo, el costo y las implicaciones en términos de tiempo de trasladar por avión a un consultor/a desde Washington D.C. para trabajar en una evaluación en la costa occidental de Estados Unidos (a unos 4,800 kilómetros) son similares a los que implicaría el traslado por avión de una persona desde Inglaterra hasta Nigeria. Del mismo modo, en muchas áreas de proyectos en Estados Unidos el inglés no constituye el idioma materno; por ello, la competencia y la sensibilidad lingüística y cultural son igualmente importantes en dichos lugares.

A continuación les presentamos algunas reglas para seleccionar y utilizar consultores/as:

Cerciórese de que las organizaciones locales y el cliente participen activamente en la definición de los requisitos para el consultor/a externo así como en el proceso de selección del mismo.

Considere cuidadosamente los méritos de un consultor/a internacional en oposición a uno/a nacional. A menudo es necesario hacer concesiones mutuas entre la mayor experticia técnica del consultor/a internacional y el conocimiento local (y desde luego la habilidad lingüística) del consultor/a nacional. Asimismo, no utilizar consultores/as nacionales puede antagonizar a la comunidad de profesionales locales, que podrían resistirse a cooperar con el experto/a internacional.

Si se contrata a un consultor/a internacional, conceda prioridad a aquellos candidatos/as que posean experiencia de trabajo en el país en cuestión y dominen el idioma local (si procede).

En el caso de evaluaciones con un enfoque operativo, evite seleccionar consultores/as con credenciales académicas impresionantes pero con experiencia de campo limitada en la ejecución de evaluaciones de

programas. Los requisitos son distintos a los de un proyecto de investigación de corte académico.

Es muy común que los consultores/as externos no se utilicen de la manera más costo-efectiva posible, ya sea porque realizan muchas de las funciones de las que podrían ocuparse igualmente bien o mejor algunos miembros del personal local o porque son incorporados en el momento menos útil. Hay aquí algunas sugerencias de opciones para garantizar un uso eficaz de los consultores/as externos:

Defina cuidadosamente qué se le está pidiendo al consultor/a que haga y considere si todas estas actividades son necesarias.

Incluso si el presupuesto le permite, intente programar el tiempo suficiente para que el consultor/a se familiarice con la organización, el proyecto y los contextos en que éste está siendo ejecutado. Un consultor/a que no comprenda el proyecto, no haya pasado cierto tiempo en las comunidades o no haya desarrollado una buena relación con el personal del proyecto, los clientes y otros interesados, será de poca utilidad.

Planee cuidadosamente en qué momentos debe ser involucrado el consultor/a y haga las gestiones necesarias con anticipación para cerciorarse de que el consultor/a se encuentre disponible en las fechas en que se le requiera. No ceda ante los consultores que desean cambiar las fechas, especialmente a corto plazo, para adecuarlas a su propia conveniencia. Algunos de los momentos cruciales para involucrar a un consultor/a son:

Durante la fase de delimitación del alcance de la evaluación, que es cuando se están tomando las decisiones más importantes relacionadas con los objetivos, el diseño y los métodos de recolección de datos y se está llegando a un acuerdo con el cliente acerca de las alternativas para hacer frente a las restricciones de tiempo, presupuesto e información;

Cuando se están tomando las decisiones relativas al tamaño y el diseño de la muestra;

Cuando se están evaluando los resultados de la ronda inicial de datos recolectados;

Cuando se está preparando el primer borrador del informe de evaluación;

Cuando se están presentando los hallazgos de la evaluación a los distintos interesados.

Encargue al personal de la institución o a consultores/as locales que preparen un documento informativo (un estudio preparatorio) antes que el consultor/a externo empiece el trabajo. El documento debe contener un resumen de información importante relativa al proyecto

(incluyendo una compilación de documentos clave), las principales instituciones asociadas y las locaciones donde se está llevando a cabo el proyecto. El documento, que debe elaborarse en coordinación con el consultor/a (por ejemplo, mediante intercambios por correo electrónico o por teléfono), podría incluir además estudios de diagnóstico rápido de unas cuantas comunidades. Un documento de esta naturaleza, bien preparado, puede representar un ahorro de tiempo considerable para el consultor/a y desencadenar un diálogo sobre los temas y prioridades clave entre los clientes, investigadores/as locales e interesados, con anterioridad a la llegada del consultor/a.

Considere el uso de videoconferencias o teleconferencias para que el consultor/a pueda mantener un contacto más asiduo con otras personas involucradas en la planificación y ejecución de la evaluación. Esto permitirá que el consultor/a colabore con etapas decisivas de la evaluación sin tener que estar siempre físicamente presente. De este modo, el consultor/a podrá hacer sugerencias sobre la muestra u otros elementos del diseño de evaluación en una fase lo suficientemente incipiente como para poder realizar cambios basados en sus recomendaciones. Las videoconferencias o teleconferencias ofrecen una ventaja adicional, la flexibilidad, al evitar situaciones extremadamente costosas en las que, por ejemplo, un consultor/a vuela desde Europa hasta África Occidental para participar en la fase de diseño de un proyecto solo para descubrir a su llegada que ésta ha sido postergada durante varias semanas.

Considere el uso de especialistas en áreas de contenido.

Además de experticia en competencias de evaluación propiamente dichas (por ejemplo, entrevistas cualitativas, elaboración de cuestionarios, diseño de la muestra y análisis de datos), es igualmente importante que al menos un miembro del equipo cuente con experiencia en el área de contenido de la evaluación (por ejemplo, extensión agrícola, educación secundaria, microcrédito). Idealmente, si los recursos así lo permiten, el equipo debe incluir entre sus miembros tanto a un experto/a sectorial con experiencia en una variedad de países o programas diferentes como a una persona con conocimiento del área local. El sistema escolar o de salud de Chicago o Dhaka probablemente tendrá una serie de características singulares (culturales, institucionales y políticas) que será necesario incorporar a la evaluación.

El Capítulo 14 argumenta lo mismo en relación con el diseño de la muestra y el análisis estadístico. La eficacia del diseño de la muestra o la aplicación de las pruebas estadísticas apropiadas a menudo se ven comprometidas por la falta de familiarización del experto/a en estadística con la literatura sobre prácticas convencionales en un área en particular (por ejemplo, pruebas de educación o una rama de la investigación psicológica).

Sea creativo en relación con los recolectores/as de datos.

En ocasiones se cuenta con opciones creativas para reducir el costo de contratación de recolectores/as de datos. En una evaluación de salud posiblemente se pueda contratar a estudiantes de enfermería; en una evaluación del sector agropecuario, a extensionistas agrícolas, así como, en muchos tipos de evaluaciones, a estudiantes universitarios/as como entrevistadores/as o encuestadores/as. Con frecuencia se puede coordinar con el hospital escuela, el ministerio de agricultura o un profesor/a universitario la contratación de estudiantes o trabajadores/as a una tarifa que sea satisfactoria para ellos/as pero inferior a los precios del mercado. Aunque las opciones mencionadas pueden ser atractivas en términos del potencial ahorro de costos que representan o por la oportunidad que ofrecen de desarrollar capacidades locales en evaluación, desde el punto de vista de la calidad albergan peligros evidentes. Los entrevistadores/as pueden no tomar en serio la asignación, o puede ser políticamente difícil seleccionar tan solo a los que parezcan más competentes o tomar medidas contra los que hayan realizado un trabajo de mala calidad. Además, los costos de supervisión y capacitación pueden ser altos y la recolección de datos puede requerir un tiempo más largo. Sin embargo, la experiencia ha demostrado que estos tipos de cooperación pueden funcionar bastante bien si existe un compromiso serio por parte de la institución o facultad universitaria.

Otra opción creativa es emplear recolectores/as de datos de la comunidad. En ocasiones una escuela secundaria local podrá llevar a cabo un estudio de evaluación de las necesidades de la comunidad, o una organización comunitaria podrá realizar estudios de base o el seguimiento de los avances de un proyecto. Otra alternativa consiste en emplear técnicas de auto-información. Por ejemplo, personas individuales o familias pueden llevar un diario de ingresos y egresos; el uso diario del tiempo; o el tiempo, el medio y el destino de los desplazamientos. Se pueden repartir cámaras fotográficas, grabadoras o cámaras de video a algunos grupos de la comunidad, encargándoles que graben ciertas situaciones, por ejemplo, los problemas que aquejan a los jóvenes, las necesidades de la comunidad o el estado de la infraestructura comunitaria. Aunque estas técnicas pueden ser cuestionadas en relación con la validez de la información, resultan útiles para recoger la perspectiva de la comunidad sobre los temas que están siendo estudiados.

6. Recolectar los datos con eficiencia

Simplifique los planes de recolección de datos.

La recolección de datos tiende a ser uno de los rubros más caros y que demandan más tiempo en una evaluación. Por consiguiente, cualquier esfuerzo para reducir los costos o el tiempo implicará casi inevitablemente simplificar los planes de recolección de datos. Para ello se requieren básicamente tres enfoques:

1. Discutir con el cliente qué información es realmente necesaria para la evaluación y eliminar cualquier otra información que figure en los términos de referencia o haya sido mencionada en discusiones subsiguientes y no resulte esencial para responder a las preguntas clave de la evaluación.

2. Revisar los instrumentos de recolección de datos con el objeto de eliminar información innecesaria. Los instrumentos de recolección de datos tienden crecer a medida que diferentes personas sugieren rubros adicionales que sería “interesante” incluir - aunque no estén directamente relacionados con el propósito de la evaluación.
3. Depurar el proceso de recolección de datos con el fin de reducir los costos y plazos. La Tabla 3 (ver también el Capítulo 3) ofrece un resumen de las distintas estrategias disponibles para reducir los costos en la fase de recolección de datos. Éstas incluyen:

- Simplificar el diseño de evaluación (por ejemplo, eliminar la recolección de datos de base o suprimir el grupo de comparación);
- Establecer las necesidades de información prioritaria del cliente (tal como vimos anteriormente);
- Buscar datos secundarios confiables (tal como vimos anteriormente);
- Reducir el tamaño de la muestra (ver el Capítulo 14);
- Reducir los costos de recolección, ingreso y análisis de datos (por ejemplo, cuestionarios autoadministrados, observaciones directas en vez de encuestas, entrevistas con grupos focales y foros comunitarios en vez de encuestas de hogares y contratación de recolectores/as de datos más económicos).

La Tabla 8 ofrece una relación de las opciones disponibles para reducir el tiempo requerido para recolectar y analizar los datos, mientras que la Tabla 9 ofrece una descripción de una serie de técnicas de recolección rápida de datos.

Encargue estudios preparatorios.

En ocasiones se puede lograr un ahorro considerable de costos y tiempo encargando a un miembro del personal de la organización o a un consultor/a local que elabore un estudio preparatorio. Éste puede abarcar los siguientes aspectos:

- Descripción de los diferentes componentes del proyecto objeto de evaluación y la forma en que están organizados;
- Información básica sobre la entidad ejecutora;
- Estudios de diagnóstico rápido de las comunidades del proyecto y posibles comunidades de comparación;
- Información sobre las instituciones estatales, **ONG** y otras organizaciones involucradas en o familiarizadas con el proyecto;
- Recomendaciones de líderes comunitarios u otros informantes clave con los cuales el consultor/a internacional debe reunirse y recopilación de información general sobre ellos/as.

Busque datos secundarios confiables.

Reunir datos secundarios confiables y pertinentes permite ahorrar una gran cantidad de tiempo y dinero. Dependiendo del país y de los temas en consideración, es posible que se pueda buscar en los archivos de los institutos de estadísticas o departamentos de planificación del gobierno, universidades u otros centros de investigación, escuelas, bancos comerciales o programas de crédito, medios de comunicación masiva y muchos sectores de la sociedad civil. De hecho, el evaluador/a debe hacer uso de cualesquier registros pertinentes, por ejemplo, de los datos resultantes del seguimiento realizado por la entidad ejecutora y los informes anuales producidos por ésta. Estos registros pueden llevarse para fines de planificación, control administrativo y financiero, evaluación de avances o comunicación con diferentes grupos cuya autorización, apoyo financiero o aprobación general son fundamentales para el éxito de la organización y sus actividades. En las secciones precedentes de este capítulo se trataron algunos de los aspectos importantes en los que es preciso incidir al evaluar las fortalezas y debilidades de diversos tipos de datos secundarios.

Ojo: Nunca acepte datos secundarios por su valor aparente, sin haber verificado antes su confiabilidad.

Si bien muchos de los problemas de los datos secundarios se refieren a las diferencias en el período de tiempo cubierto, la cobertura inadecuada de algunos sectores de la población objetivo o la mala calidad de la recolección de datos y de los informes sobre los mismos, en ocasiones los datos recogidos acusan debilidades más importantes. Cuando la calidad de la supervisión es deficiente, un número significativo de encuestas puede ser enteramente falsificado por los entrevistadores/as, o secciones importantes de las encuestas pueden ser ignoradas o registradas de manera deficiente.

Recolecte tan solo los datos necesarios.

Es importante cerciorarse de recolectar información esencial solamente. La recolección de información innecesaria aumenta los costos y el tiempo y además reduce la calidad de la información requerida al fatigar a los entrevistados si tienen que responder una gran cantidad de preguntas. Por consiguiente, recomendamos revisar cuidadosamente todos los instrumentos de recolección de datos con el fin de eliminar la información que no sea pertinente y esencial para los propósitos de la evaluación y que nunca será analizada o utilizada.

Del mismo modo, es importante revisar el análisis de los datos con el fin de determinar qué tipos de análisis de datos desagregados se requieren realmente. Si se determina que ciertos tipos de desglose propuestos no son necesarios (por ejemplo, comparar los impactos del proyecto en participantes de distintas localidades), probablemente se pueda reducir el tamaño de la muestra.

Busque formas simples de recolectar datos sobre temas sensibles y de poblaciones difíciles de acceder.

Otro reto para los evaluadores/as, aunque no solo para una EMR, tiene que ver con la recolección de datos sobre temas sensibles – por ejemplo, violencia doméstica, uso de anticonceptivos o violencia en los adolescentes - o de sectores difíciles de acceder – por ejemplo, trabajadoras/es sexuales comerciales, consumidores de drogas, minorías

étnicas, poblaciones migrantes, los sin hogar o incluso, en algunas culturas, las mujeres. Hay una serie de metodologías que sirven para abordar dichos temas y acceder a los sectores mencionados. Sin embargo, las restricciones que enfrenta una EMR en términos de presupuesto, tiempo o prejuicios políticos podrían presionar para que se ignoren dichos temas sensibles o se deje de lado a los grupos de personas que son difíciles de acceder. Existen al menos tres estrategias para ayudar con el abordaje de temas sensibles:

1. Identifique una amplia gama de informantes, que puedan ofrecer perspectivas distintas.
2. Seleccione una serie de estrategias culturalmente apropiadas para investigar los temas sensibles.
3. Aplique el método de triangulación sistemáticamente.

Algunas de las metodologías culturalmente apropiadas que pueden emplearse son:

Observación participante

Observación no participante (observación de personas o grupos como persona externa, sin participar en sus actividades)

Entrevistas con grupos focales

Estudios de caso

Informantes clave

Técnicas de DRP

Los grupos difíciles de acceder incluyen: trabajadoras/es sexuales comerciales, consumidores de drogas o alcohol, delincuentes, pequeñas empresas informales no registradas, invasores de terrenos o viviendas y residentes ilegales, minorías étnicas o religiosas, parejas o padres ausentes, trabajadores/as cautivos y esclavos, vendedores/as de agua informales, niñas que asisten a escuelas de varones, trabajadores/as migrantes y personas con VIH/SIDA. En este último grupo se considera especialmente a los portadores/as del virus que no se han sometido a pruebas de descarté.

El evaluador/a puede enfrentar uno de dos escenarios. En el primer escenario la existencia de los grupos es conocida pero sus miembros son difíciles de ubicar y acceder. En el segundo escenario, los clientes y, al menos inicialmente, el evaluador/a ni siquiera son conscientes de la existencia de dichos grupos marginados o “invisibles”. Las técnicas para identificar e investigar a los grupos difíciles de acceder son similares a las que se emplean para el abordaje de temas sensibles e incluyen:

Observación participante. Se trata de una de las metodologías más comunes para familiarizarse con, y ser aceptado en, el medio en el que se desenvuelven, o se supone que se desenvuelven, los grupos. Muchas veces los contactos o presentaciones iniciales se dan a través de amigos/as, familiares, clientes o, en algunos casos, las organizaciones oficiales con las que interactúan los grupos mencionados.

Informantes clave. Entrevistas programadas con personas que están especialmente familiarizadas con, e informadas sobre, los grupos objetivo.

Estudios de seguimiento. Se recurre a vecinos, parientes, amigos, colegas del trabajo y así sucesivamente para ayudar a localizar a las personas que se han mudado.

Muestreos “bola de nieve”. En esta técnica se hace un esfuerzo para localizar a unos cuantos miembros del grupo difícil de acceder por cualesquier medios disponibles. Luego, se solicita a dichos miembros que identifiquen a otros miembros del grupo, de manera tal que, si el enfoque tiene éxito, aumentará el tamaño de la muestra. Esta técnica se utiliza con frecuencia en el estudio de problemas relacionados con enfermedades de transmisión sexual.

Técnicas sociométricas. Se pide a los encuestados que identifiquen a quién recurren en busca de consejo o ayuda en relación con temas específicos (por ejemplo, asesoría sobre planificación familiar, medicina tradicional o la compra de sustancias ilegales). Acto seguido se traza un mapa sociométrico, con flechas que conectan a los informantes con líderes de opinión, informantes o especialistas.

7. Analizar los datos con eficiencia

Busque formas de administrar los datos con eficiencia.

Antes de analizar los datos, es necesario ingresarlos en un formato electrónico o manual. No hacerlo de la manera apropiada puede comprometer la calidad y confiabilidad de los datos u originar pérdidas de tiempo, dinero o ambos. Más aún, si no se administran correctamente los datos, se corre el riesgo de perder cantidades significativas de información. Los siguientes pasos son importantes para la elaboración y ejecución de un plan de análisis:

Elaborar un plan de análisis (ver Tabla 11.4*, Capítulo 11), especificando, para cada tipo de análisis propuesto, los objetivos del análisis, la hipótesis a ser validada, las variables incluidas en el análisis y los tipos de análisis a ser efectuados.

Elaborar y validar el libro de códigos. Si el cuestionario contiene preguntas abiertas, es necesario revisar las respuestas para definir las categorías que se usarán. Si se ha clasificado algunos de los datos numéricos por categorías (“más de una vez por semana”, “una vez por semana”, etc.), deberán revisarse las respuestas para identificar posibles problemas o inconsistencias.

Garantizar una codificación confiable. Es decir, cerciorarse de que el libro de códigos esté completo y posea coherencia lógica; además, monitorear el proceso de codificación de los datos para verificar la exactitud y la coherencia entre codificadores.

Revisar las encuestas en búsqueda de datos faltantes y decidir qué tratamiento dar a los datos faltantes (ver el Capítulo 14). En algunos casos será posible regresar al campo o enviar nuevamente los cuestionarios por correo a los encuestados, pero en la mayoría de los casos estas medidas resultan poco prácticas. Los datos faltantes muchas veces no son azarosos; en consecuencia, el tratamiento que se les dé es importante para evitar los sesgos. Por ejemplo, puede haber diferencias entre los sexos, grupos de edad y grupos económicos o educativos en lo que respecta a la disposición de los mismos para responder ciertas preguntas. Además, puede haber diferencias entre grupos étnicos o religiosos, o entre propietarios e invasores de tierras. Uno de los primeros pasos del análisis debe ser preparar distribuciones de frecuencias de los datos faltantes, a fin de determinar variables clave y, ahí donde proceda, efectuar un análisis exploratorio para establecer si existen diferencias significativas en las tasas de datos faltantes entre los grupos poblacionales clave antes mencionados.

Ingresar los datos en el sistema de análisis de datos manual o computarizado.

Depurar los datos. Este proceso supone lo siguiente:

Realizar un análisis de datos exploratorio con el fin de identificar los datos faltantes así como potenciales problemas, por ejemplo, variables con valores fuera de los rangos esperados. Se trata de variables en las que unos cuantos valores de una variable en particular acusan rangos muy por encima o muy por debajo del rango normal. Unas cuantas variables de este tipo pueden afectar el análisis significativamente, dificultando enormemente la obtención de resultados estadísticamente significativos (dado que la desviación estándar se incrementa dramáticamente). En consecuencia, el proceso de depuración de datos debe incluir reglas claras sobre el tratamiento que debe darse a dichas variables (ver el Capítulo 11).

Decidir cómo tratar los datos faltantes y qué políticas se aplicarán al respecto.

Identificar variables que pudieran requerir recodificación.

*Documentar detalladamente cómo se depuraron los datos, qué tratamiento se le dio a los datos faltantes y cómo se crearon los índices.*⁵

⁵ En ocasiones se crea un índice mediante una combinación de varias variables, o por transformación estadística. Por ejemplo, se puede crear la variable “consumo” para representar la suma de gastos en alimentos, vivienda, transporte, educación y otros rubros. Es fundamental que los procedimientos

Michael Bamberger, Jim Rugh y Linda Mabry *RealWorld Evaluation: Working under Budget, Time, Data and Political Constraints*. ©Sage Publications, 2006. Reimprimir con autorización previa de los autores solamente. Favor contactar a: jmichaelbamberger@gmail.com o JimRugh@MindSpring.com.

Aunque la EMR se ciñe a la mayoría de los procedimientos estándar del análisis de datos, cuando se enfrentan restricciones de tiempo o presupuesto es posible que se tenga que aplicar una serie de enfoques. Cuando la restricción principal es el *tiempo* y existen recursos adicionales disponibles para acelerar el proceso, se pueden considerar los siguientes enfoques:

- Ingreso directo de los datos de la encuesta en computadoras portátiles
- Uso del escaneo electrónico para leer los cuestionarios
- Subcontratación del análisis de datos a una universidad o una organización de investigación de mercado
- Contratación de más, o más experimentados, codificadores y analistas de datos

Cuando la restricción principal es el *dinero*, se pueden considerar una o más de las siguientes opciones:

- Limitar los tipos de análisis estadístico a fin de reducir el siempre caro “tiempo de computación”
- Considerar la adquisición y el uso de paquetes estadísticos populares como el SPSS o el SAS, de manera tal que el análisis pueda ser efectuado dentro de la propia organización en vez de subcontratar a terceros. Desde luego, para esta opción se requiere la disponibilidad de expertos/as en estadística al interior de la organización.

Focalice el análisis en responder las preguntas clave.

En cualquier evaluación, un consejo inteligente es concentrarse en las preguntas clave, es decir, en aquellas que se relacionan con el propósito principal de la evaluación. Esto es particularmente importante en el caso de una EMR, donde se tienen que tomar decisiones respecto de qué cosas eliminar debido a las limitaciones de tiempo y financiamiento. Ser recordados de cuáles son las preguntas principales y qué se requiere para responderlas adecuadamente ayudará a quienes están planificando una EMR a concentrarse en dichos temas, y no en otros. Típicamente, los clientes e interesados, así como los propios evaluadores/as, quisieran recolectar información adicional. Sin embargo, al estar sujetos a las restricciones de una EMR, lo que sería “interesante de averiguar” debe ser separado de lo que “resulta esencial”, a fin de responder las preguntas clave que constituyen la razón de ser de la evaluación.

Como vimos en el Capítulo 2, y nuevamente en secciones anteriores de este capítulo, algunos ejemplos de preguntas clave típicas de una evaluación incluyen:

- ¿Existen evidencias de que el proyecto ha cumplido (o está en camino de cumplir) sus objetivos? ¿Se han producido cambios verificables en las

para crear todos los índices estén claramente documentados y sean fácilmente accesibles para los lectores/as.

Michael Bamberger, Jim Rugh y Linda Mabry *RealWorld Evaluation: Working under Budget, Time, Data and Political Constraints*. ©Sage Publications, 2006. Reimprimir con autorización previa de los autores solamente. Favor contactar a: jmichaelbamberger@gmail.com o JimRugh@MindSpring.com.

- características de la población objetivo en relación con los impactos que el proyecto estaba tratando de generar? ¿Qué objetivos fueron (y no fueron) alcanzados? ¿Por qué? ¿Es razonable asumir que los cambios se debieron al proyecto, y no a factores externos (no controlados por los ejecutores del proyecto)?
- ¿El proyecto se planteó los objetivos correctos? ¿Se basaron éstos en un diagnóstico adecuado de las causas subyacentes del problema o problemas a ser abordado(s)?
- ¿Qué impacto ha tenido el proyecto en diferentes sectores de la población objetivo – incluyendo los sectores más pobres y vulnerables? ¿Son diferentes los impactos en los hombres y las mujeres? ¿Hay grupos étnicos, religiosos u otros grupos similares que no se hayan beneficiado o hayan sido afectados negativamente por el proyecto?
- ¿Los efectos directos son sostenibles y los beneficios tienen probabilidades de sostenerse en el tiempo? ¿Las comunidades o los grupos objetivo son razonablemente representativos de la población más amplia (por ejemplo, todos los agricultores pobres o todos los pobladores de zonas urbano-marginales) y qué probabilidades existen de lograr los mismos impactos si el proyecto es replicado a mayor escala?
- ¿Cuáles son los factores contextuales y externos que determinan el grado de éxito o fracaso del proyecto?

El evaluador/a en un mundo real debe establecer cuáles son los temas clave que necesitan ser explorados en profundidad y qué temas son menos gravitantes y pueden ser estudiados con menos intensidad, o eliminados por completo. Es imprescindible, además, determinar en qué ocasiones el cliente necesita un análisis estadístico riguroso (y caro) para legitimar los hallazgos de la evaluación frente a los miembros del congreso o parlamento, o ante organizaciones financieras críticas del programa, y en qué oportunidades sería aceptable producir un análisis y hallazgos más generales. La respuesta a estas preguntas puede tener un impacto considerable en el presupuesto de la evaluación y en los plazos requeridos, especialmente en lo que se refiere al diseño y al tamaño de la muestra.

8. Informar sobre los hallazgos de manera eficiente y efectiva

Como mencionamos en la sección anterior, titulada “Adaptar los planes de evaluación”, una evaluación debe focalizarse en las preguntas clave vinculadas al propósito principal por el cual se está ejecutando la evaluación. Esto es especialmente importante en el caso de una EMR, donde se tienen que tomar decisiones sobre lo que puede y no puede ser eliminado debido a las limitaciones cronológicas y financieras. Es necesario tener presentes las preguntas clave no solamente durante la planificación de la evaluación o la recolección y el análisis de los datos, sino también durante la redacción del informe o informes de evaluación. Existe la tentación de informar sobre todo tipo de

“hallazgos interesantes”, pero el evaluador/a o evaluadores/as deben mantener el informe enfocado en responder las preguntas clave que el cliente o clientes y los interesados desean ver respondidas.

Una de las opciones más eficaces para aumentar la probabilidad de que los hallazgos de una evaluación sean efectivamente utilizados es cerciorarse de que revistan utilidad práctica directa para los diferentes interesados. Algunos de los factores que afectan el uso de los hallazgos son:

La oportunidad en que se realiza la evaluación

Reconocer que la evaluación es solo una entre varias fuentes de información e influencia para quienes toman las decisiones, y cerciorarse que la evaluación sea un complemento para las otras fuentes

Desarrollar una relación continua con interesados clave, estando cuidadosamente atentos a sus necesidades, estableciendo qué opinan del contexto político y manteniéndolos informados de los avances de la evaluación. No debe haber “sorpresas” al presentar el informe de evaluación (Departamento de Evaluación de Operaciones, 2005; Patton, 1997).

Algunos pasos importantes inherentes a la presentación de los hallazgos de una evaluación incluyen:

Conozca a los interesados de la evaluación y establezca de qué forma les gusta recibir información.

Apóyese en una presentación visual para complementar los informes escritos o presentaciones orales. Ahí donde sea apropiado y viable, utilice herramientas de presentación - por ejemplo, PowerPoint - pero no se convierta en un esclavo de la tecnología y esté preparado para trabajar sin dichos sistemas si la logística se vuelve demasiado complicada. Las presentaciones visuales son especialmente útiles cuando la presentación no se lleva a cabo en el idioma materno de gran parte de los miembros de la audiencia.

Comparta los resultados de la evaluación por medio de presentaciones orales. Muchos interesados no se sienten cómodos con los informes escritos o con presentaciones apoyadas en diapositivas; por ello, exponer los hallazgos verbalmente puede ser importante.

Planifique el informe escrito de manera tal que sea sencillo, atractivo y amable con el usuario. Considere presentar varias versiones de los hallazgos, en los formatos que resulten más comprensibles y útiles para diferentes destinatarios. (Más detalles sobre este punto más adelante.)

Involucre a los medios de comunicación masiva. Cuando uno de los fines de la evaluación es llegar a, e influir en, un público más amplio (por ejemplo, la opinión pública en general, todos los padres y madres de

niños y niñas en edad escolar secundaria, el poder legislativo), la prensa puede ser un valioso aliado. Sin embargo, trabajar con los medios requiere tiempo y preparación y, si su participación es importante, es posible que valga la pena contratar a un consultor/a que “conozca los entretelones”.

Informe sucinto a los clientes primarios

Cuando los hallazgos y recomendaciones no llegan a los clientes primarios en un formato que les sea agradable y entendible, esto reduce el impacto de muchas evaluaciones. No existe un único formato que se considere óptimo para informar sobre los hallazgos de una evaluación; la decisión sobre el formato a utilizar dependerá más bien de los clientes y la naturaleza de la evaluación. Un buen punto de partida es preguntar a los clientes qué informes previos les resultaron especialmente útiles y por qué.

Una norma general, en especial para una EMR en la que el tiempo constituye una restricción, es hacer una presentación corta y sucinta. Es aconsejable disponer de un documento físicamente breve que pueda ser ampliamente distribuido; aunque el resumen ejecutivo al inicio de un informe largo esté muy bien escrito, algunos clientes e interesados se sentirán intimidados por el tamaño del documento y pueden inhibirse de leer siquiera el resumen.

Vaughan y Buss (1998) proponen algunos parámetros valiosos para decidir qué información transmitirles a los ocupados diseñadores de políticas y cómo transmitirlo. Señalan que muchos diseñadores de políticas poseen la capacidad intelectual para leer y entender análisis complicados pero que la mayoría no tiene tiempo. En consecuencia, muchos desearán que se les transmita una idea general de las complejidades del análisis (no desean ser tratados de manera condescendiente), pero sin extraviarse en los detalles. Otros diseñadores de políticas carecerán de la formación técnica necesaria y desearán una presentación más bien sencilla. De modo que debemos establecer un balance delicado entre mantener el respeto y el interés de los destinatarios más técnicos sin perder el de los menos técnicos. Lo que sí es seguro, sin embargo, es que nadie tiene tiempo. Por consiguiente, la presentación debe ser corta, aunque no necesariamente simple. Las reglas de Vaughan y Buss para decidir qué información incluir son las siguientes:

Analice las políticas pero no la política. Los evaluadores/as son contratados para proporcionar experticia técnica, no para fungir de asesores sobre estrategia política.

Mantenga el formato simple.

Comunique los aspectos fundamentales pero también el razonamiento.

Muchos diseñadores de políticas querrán saber cómo llegó el evaluador/a a las conclusiones del informe, para poder sopesar cuánto peso darle a los hallazgos.

Utilice cifras en pequeñas dosis.

Esclarezca, no defienda. Si los evaluadores/as defienden una política en particular, arriesgan perder la confianza del diseñador de políticas.

Identifique ganadores y perdedores. Quienes toman las decisiones se interesan por la forma en que las políticas afectarán su base de apoyo, especialmente a corto plazo. En consecuencia, si los evaluadores/as y analistas desean que los diseñadores de políticas los escuchen, deberán identificar ganadores y perdedores. Por ejemplo, uno de los argumentos más eficaces del estudio sobre las razones para cerrar el sumamente caro pero políticamente sensible programa de raciones subsidiadas de harina de trigo en Pakistán fue el análisis de quiénes serían los potenciales perdedores (los distribuidores de harina de trigo y los propietarios de negocios minoristas) y cómo podrían mitigarse sus pérdidas (Departamento de Evaluación de Operaciones, 2005, Capítulo 6).

No omita mencionar las consecuencias no intencionales. Las personas a menudo responden a políticas y programas nuevos de maneras inesperadas, especialmente cuando se trata de tomar ventaja de nuevos recursos u oportunidades. En algunas ocasiones, las reacciones inesperadas pueden destruir un programa potencialmente bueno, mientras que, en otras, los efectos directos no anticipados de un programa pueden añadir al éxito del mismo. Quienes diseñan las políticas son sensibles a las consecuencias imprevistas porque entienden que tienen un costo político o económico potencialmente alto. Consecuentemente, si la evaluación puede identificar algunas consecuencias importantes de las cuales no habían tomado conciencia los diseñadores de políticas, concitarán la atención del público objetivo y elevarán la credibilidad de la evaluación.

Informes prácticos, entendibles y útiles para otros destinatarios

Adicionalmente al cliente y otros interesados primarios (por ejemplo, los ministerios de estado concernidos y la institución financiera), a menudo existirán otros interesados que tengan interés en la evaluación por diferentes motivos. Algunos sectores, por ejemplo la población objetivo, se verán directamente afectados por la evaluación; otros participarán en acciones de incidencia y desearán utilizar los hallazgos ya sea para respaldar sus argumentos o para criticar el informe porque no corrobora sus posiciones; y nuevamente otros se interesarán por la aplicación práctica de los hallazgos. Muchas veces el cliente no desea que se difundan muy ampliamente los hallazgos de la evaluación, especialmente si éstos son críticos y pueden sacar a la luz algunos temas sensibles. En estos casos, los evaluadores/as posiblemente tengan que plantearse consideraciones éticas y profesionales delicadas, en el sentido de preguntarse si tienen la obligación ética, y quizás profesional, de difundir los hallazgos de la evaluación a todos los sectores afectados por el proyecto, contradiciendo las instrucciones del cliente de limitar la distribución. Estos aspectos, de gran trascendencia, se analizan en los Capítulos 6 y 7.

Asumiendo que los referidos aspectos éticos se resuelvan satisfactoriamente, es necesario definir una estrategia de difusión para llegar a sectores con diferentes campos de interés, niveles de experticia en la lectura de informes de evaluación y preferencias en términos de cómo les gusta recibir la información. En algunos casos, es posible que

algunos grupos deban recibir el informe además en otro(s) idioma(s). El equipo de evaluación debe decidir qué interesados son lo suficientemente importantes como para justificar la preparación de una versión distinta del informe (y quizás incluso su traducción a otro idioma) o la organización de presentaciones y discusiones por separado.

Estas consideraciones son especialmente importantes para las EMR porque llegar a los diferentes destinatarios, especialmente los sectores más pobres, menos educados y menos accesibles, tiene implicaciones significativas en términos de costo y tiempo. Existe el peligro de que, cuando se enfrentan restricciones presupuestarias o cronológicas, la evaluación llegue únicamente a los principales clientes, y muchos de los grupos cuyas vidas son las más afectadas (por ejemplo, grupos indígenas cuyo estilo de vida se encuentra amenazado, invasores urbanos que son reubicados por el uso de la fuerza o comunidades de bajos ingresos que pueden beneficiarse o no de una nueva tecnología de agua y saneamiento) posiblemente nunca vean la evaluación ni sean consultados jamás respecto de sus conclusiones y recomendaciones.

Un propósito importante del ejercicio de delimitación del alcance de la evaluación (el primer paso del enfoque de EMR) es acordar con el cliente quién recibirá el informe de evaluación y tendrá la oportunidad de expresar sus puntos de vista acerca del mismo. Si el cliente demuestra escaso interés en una difusión amplia, pero no se opone activamente a ella, en ese caso el evaluador/a puede proponer estrategias costo-efectivas para llegar a un público más amplio. Si, por el contrario, el cliente se opone activamente a un proceso de consulta o difusión más amplio, en ese caso el evaluador/a deberá considerar las opciones – una de las cuales consistiría en declinar aceptar el contrato de evaluación.

Si las principales restricciones a una difusión más amplia son de tiempo y presupuesto, las opciones disponibles serían las siguientes:

- Conseguir el apoyo de los medios de comunicación masiva. A menudo se necesita invertir un tiempo considerable en cultivar relaciones con periodistas de la televisión, radio y prensa escrita. Se les puede invitar a acompañar a los evaluadores/as en sus visitas de campo o reuniones con la comunidad y se les puede enviar cada cierto tiempo historias interesantes que sirvan de material para noticias o reportajes.

- Conseguir el apoyo de las ONG y organizaciones de la sociedad civil. Éstas a menudo estarán dispuestas a colaborar con la difusión de los hallazgos pero es posible que deseen presentarlos bajo su propia óptica (que puede ser muy diferente a la del equipo de evaluación); por eso, es importante familiarizarse con las distintas organizaciones antes de invitarlas a colaborar con la difusión.

- Organizar reuniones con organizaciones de las comunidades objetivo para presentarles los hallazgos y obtener retroalimentación. Es importante organizar estas reuniones con la debida antelación durante el proceso de preparación del informe, de modo que las opiniones e informaciones recabadas puedan ser incorporadas al informe final.

9. Ayudar a los clientes a utilizar los hallazgos correctamente

Lamentablemente, es muy común llevar a cabo una evaluación, escribir el informe formal, entregárselo al cliente y que luego no suceda nada más. Si se tienen en cuenta las recomendaciones anteriores, que incluyen involucrar al cliente y otros interesados clave a todo lo largo del proceso de evaluación, uno esperaría que los hallazgos de una evaluación fuesen relevantes y se tomaran en serio. Sin embargo, si no se hace un seguimiento ex post podría parecer que la evaluación no hubiese tenido ningún valor. Algunas organizaciones financieras importantes, tras advertir el uso limitado de los informes de evaluación, han decidido simplemente dejar de encargar evaluaciones rutinarias. ¿No sería mejor invertir un esfuerzo mayor en cerciorarse de que las evaluaciones se focalicen en responder las preguntas clave, se ejecuten bien y luego sean mejor aprovechadas?

Un importante propósito de la EMR es ayudar a los involucrados a enfocarse en los aspectos más importantes y maximizar la eficiencia en la ejecución de evaluaciones que añadan valor y sean útiles. El paso final – la utilización – debe formar parte de esta fórmula de eficiencia. Si la información no se utiliza para enriquecer decisiones que lleven a una mejor calidad y eficacia de los programas, la evaluación habrá sido un desperdicio. El argumento aquí es que quienes realizan una evaluación tienen que tener en cuenta que preocuparse por el aspecto de la utilización es un componente importante del proceso de evaluación.

Una forma de hacerlo es ayudando al cliente a elaborar un plan de acción que especifique los pasos que se van a tomar en respuesta a las recomendaciones de la evaluación, y luego hacer un seguimiento de la ejecución del referido plan de acción. Hacer esto es evidente si se trata de una **evaluación formativa**, cuyos hallazgos se utilizan para mejorar la ejecución subsiguiente de un proyecto en marcha. Incluso en el caso de una evaluación sumativa (cuyo propósito es estimar la medida en que se han alcanzado los efectos directos y los impactos de un proyecto), o de la evaluación de un proyecto que ya concluyó, el seguimiento ex post debe incluir ayudar a utilizar las lecciones aprendidas para enriquecer estrategias futuras así como el diseño de futuros proyectos. Como mínimo, los responsables de una evaluación tienen que hacer todo lo posible para cerciorarse de que los hallazgos y recomendaciones sean documentados y comunicados en formatos útiles a quienes toman las decisiones tanto ahora como en el futuro.

Lecturas adicionales

Al final de este capítulo hemos incluido una breve relación de lecturas adicionales. Se recomienda a los lectores/as revisar la lista de lecturas al final de cada capítulo del libro, así como las extensas referencias al final del libro. El Anexo 4 contiene asimismo una relación de recursos electrónicos sobre muchos de los temas abordados en este capítulo.

Tabla 1: Algunas de las formas en que las influencias políticas afectan las evaluaciones	
Durante el diseño de la evaluación	
Criterios de selección de los evaluadores/as	<p>Los evaluadores/as pueden ser seleccionados con base en:</p> <ul style="list-style-type: none"> • su imparcialidad o experticia profesional • su simpatía por el programa • sus críticas conocidas del programa (cuando el cliente desea utilizar la evaluación para recortar el programa) • la facilidad con la que pueden ser controlados • su nacionalidad coincidente con la del país o estado de la institución financiera del programa
Selección del diseño de evaluación y de los métodos de recolección de datos	<p>La decisión de utilizar un enfoque ya sea cuantitativo o cualitativo, o de recolectar datos que puedan ser ingresados en un tipo de modelo analítico determinado (por ejemplo, recolectar datos econométricos o sobre el rendimiento estudiantil en un programa de educación), puede predeterminar los temas en los que incidirá y no incidirá la evaluación.</p> <p>El uso de grupos de control puede ser descartado por razones éticas o políticas más que metodológicas, por ejemplo:</p> <ul style="list-style-type: none"> • para evitar generar expectativas de compensación entre los grupos de control • para no tener que negar beneficios a algunos sectores de la comunidad que los necesitan • para evitar presiones de expandir el proyecto a las áreas de control • para evitar incidir en grupos políticamente sensibles o volátiles <p>Por otra parte, los evaluadores/as pueden insistir en incluir grupos de control en el diseño de evaluación porque desean ceñirse a las prácticas convencionales de su profesión, incluso si dichos grupos contribuyen poco al abordaje de las preguntas de evaluación.</p>
<i>Ejemplo de una decisión específica relacionada con el diseño:</i> Utilizar o no grupos de control (es decir, un diseño cuasi-experimental)	
Selección de indicadores e instrumentos	<p>La decisión de utilizar indicadores cuantitativos solamente puede llevar (intencionalmente o no) a ciertos tipos de hallazgos y excluir el análisis de otros temas potencialmente sensibles. Por ejemplo, si solo se utilizan cuestionarios estructurados probablemente no se mencionarán los problemas de violencia doméstica o acoso sexual en el transporte público.</p>
Selección de los interesados que serán involucrados o consultados	<p>Si solo se consulta a los funcionarios/as del gobierno, el diseño de la evaluación y los temas abordados pueden ser bastante diferentes, por ejemplo, a los de una evaluación del mismo programa en la cual se consulta también a las organizaciones comunitarias, los jefes/as de hogar y las ONG. El evaluador/a puede ser disuadido, ya sea formal o informalmente, de recolectar información de determinados sectores sensibles, por ejemplo, limitando el tiempo o el presupuesto disponibles, un medio sutil para excluir a los grupos difíciles de acceder.</p>

Orientación profesional de los evaluadores/as	La selección de, por ejemplo, economistas, sociólogos/as, politólogos/as o antropólogos/as para llevar a cabo una evaluación tendrá un impacto importante en el diseño y los efectos directos de la misma.
Decisión sobre evaluación interna o externa	Las evaluaciones que son ejecutadas internamente, por el personal del proyecto o la organización, poseen una dinámica política distinta y están sujetas a presiones políticas distintas que las evaluaciones realizadas por consultores/as externos, en general considerados más independientes. El uso de evaluadores/as nacionales o internacionales afecta también la dinámica de la evaluación. Por ejemplo, si bien es cierto que los evaluadores/as nacionales tienden a conocer mejor la historia y el contexto del programa, pueden estar menos dispuestos a ser muy críticos de programas administrados por clientes regulares.
Asignaciones presupuestarias y de tiempo	Aunque las restricciones presupuestarias y cronológicas se encuentran fuera del control absoluto de algunos clientes, otros pueden intentar restringir los plazos y los recursos con el objeto de desalentar el abordaje de ciertos temas o prevenir un análisis crítico exhaustivo.
Durante la ejecución	
El rol cambiante del evaluador/a	El evaluador/a puede tener que negociar entre ejercer los roles de guía, publicista, abogado, confidente, juez severo y amigo crítico.
Selección de los destinatarios de los informes de avances y hallazgos iniciales	Un medio sutil con el que cuenta el cliente para evitar las críticas es excluir a potenciales críticos de la lista de distribución de los informes de avances. La distribución exclusiva a directores/as y coordinadores/as, excluyendo al personal del programa, o a ingenieros y arquitectos, excluyendo a trabajadores sociales y extensionistas, determinará la naturaleza de los hallazgos y los tipos de retroalimentación a los cuales se expondrá la evaluación.
Evolución de la dinámica social	Las relaciones al inicio de una evaluación suelen ser cordiales, pero pueden agriarse rápidamente cuando empiezan a emerger hallazgos negativos, o cuando el evaluador/a no sigue los consejos del cliente sobre cómo llevar a cabo la evaluación (por ejemplo, de qué grupos recolectar la información).
Difusión y uso	
Selección de los revisores/as	Si se solicita que revisen la evaluación únicamente personas que tienen un interés en la continuidad del proyecto, la retroalimentación propenderá a ser más positiva que si se involucra a críticos conocidos. Los plazos cortos, inocentes o no, pueden no conceder a algunos grupos el tiempo suficiente como para realizar comentarios significativos o incorporar sus comentarios, introduciendo un sesgo sistemático contra dichos grupos.
Selección del idioma	En los países en desarrollo son pocos los informes de evaluación que se traducen a idiomas locales, salvo cuando se trata de interesados importantes. El argumento que suele citarse como excusa es la falta de presupuesto, lo cual sugiere que informar a los interesados no es algo que el cliente considere valioso o necesario. El idioma representa un problema incluso en Estados Unidos, Canadá y Europa, donde muchas evaluaciones conciernen a las poblaciones inmigrantes.

Distribución del informe	Frecuentemente, una manera efectiva de evitar las críticas es no compartir el informe con personas críticas. Aquí puede estar en juego el interés público, por ejemplo cuando los clientes tienen una visión estrecha preestablecida de cómo deben difundirse o emplearse los resultados de la evaluación y se niegan a considerar otros usos posibles
--------------------------	--

Fuente: *RealWorld Evaluation*, Tabla 6.1.

Tabla 2: Cinco estrategias de evaluación y los diseños correspondientes

Estrategia de evaluación	Robustez metodológica del diseño de evaluación [ver Tabla 3 para una descripción de los diseños]					
	Muy robusta	Robusta	Razonable	Débil	Muy débil	Ejemplo
1. Diseño auténticamente experimental: <i>Asignación aleatoria de sujetos y estricto control del contexto dentro del cual se implementa el proyecto</i>	Diseño 1					Probar una nueva droga en condiciones de laboratorio
2. Diseño de campo aleatorio: <i>Asignación aleatoria de sujetos pero solo un control limitado del contexto del proyecto</i>		Diseño 1				Seleccionar por sorteo aldeas que participarán en un proyecto de autoayuda en el sector de agua cuando la demanda excede la oferta
3. Diseño no aleatorio robusto (cuasi-experimental): <i>Grupos del proyecto y de control pre-prueba y post-prueba</i>			Diseños 1 y 2			Un proyecto de viviendas de interés social donde se entrevista a los participantes del proyecto y grupos de comparación (de los tipos de comunidades donde antes vivían los participantes) al inicio y al final (cinco años después) del proyecto.
4. Diseños no aleatorios débiles: <i>Grupo de base o de comparación eliminado</i>				Diseños 4 y 5		Comparación post-prueba de comunidades donde se han construido carreteras rurales con comunidades similares que carecen de carreteras
5. Diseños no experimentales [solo grupo del proyecto con o sin línea de base]: <i>No se utiliza un grupo de control, lo cual dificulta establecer un contrafactual sólido desde el punto de vista lógico</i>					Diseños 6 y 7	Análisis de comunidades que cuentan con centros de salud operativos. Algunos diseños incluyen línea de base pero no hay grupo de control

Tabla 3 Los siete diseños de evaluación del impacto más ampliamente utilizados

<p>Clave T₁, T₂, etc = Etapa del proyecto (tiempo) P = Participantes del proyecto C = Grupo de control/comparación P₁, P₂, C₁, C₂, etc. = Primera, segunda (tercera y cuarta) observación de los grupos del proyecto o de control X = Intervención del proyecto (normalmente un proceso a través del tiempo antes que un evento discreto)</p>	<p>Inicio del proyecto [pre-prueba]</p>	<p>Intervención del proyecto (continúa hasta el final del proyecto)</p>	<p>Evaluación de medio término o varias observaciones durante la ejecución del proyecto</p>	<p>Final del proyecto [post-prueba]</p>	<p>Seguimiento después que el proyecto ha estado funcionando durante algún tiempo (ex post)</p>	<p>Etapa del ciclo del proyecto en la que se puede empezar a utilizar cada diseño de evaluación</p>
<p>Diseño de evaluación</p>	<p>T₁</p>		<p>T₂</p>	<p>T₃</p>	<p>T₄</p>	
<p>LOS DOS DISEÑOS DE EVALUACIÓN MÁS ROBUSTOS</p>						
<p>1. <i>Diseño longitudinal, incluyendo observaciones post-prueba y ex-post.</i> [El diseño de evaluación más comprensivo, pero su uso está limitado por requerimientos de dinero y tiempo.] Observación al inicio, durante y al final del proyecto, así como una vez concluido el proyecto. Permite evaluar la ejecución del proyecto y además observar los procesos de cambio. Rara vez es posible la asignación aleatoria a grupos del proyecto y de control; por consiguiente, tanto éste como todos los demás diseños suelen utilizar grupos de control no equivalentes.</p>	<p>P₁ C₁</p>	<p>X</p>	<p>P₂ C₂</p>	<p>P₃ C₃</p>	<p>P₄ C₄</p>	<p>Inicio o</p>
<p>2. <i>Comparación pre-prueba / post-prueba de grupos del proyecto y de control.</i> Se trata del mejor diseño práctico para casi todos los fines cuando la evaluación tiene la opción de empezar al inicio del proyecto y se dispone de un presupuesto razonable y ninguna restricción al uso de grupos de control o al acceso a datos.</p>	<p>P₁ C₁</p>	<p>X</p>		<p>P₂ C₂</p>		<p>Inicio o</p>
<p>TRES DISEÑOS DE EVALUACIÓN MENOS ROBUSTOS PERO CON FRECUENCIA ADECUADOS</p>						
<p>3. <i>Diseño longitudinal truncado.</i> Se realizan observaciones de grupos del proyecto y de control en diversos momentos durante la ejecución del proyecto pero la evaluación no empieza sino hasta que el proyecto está en marcha y por lo tanto no se dispone de información de base.</p>		<p>X</p>	<p>P₁ C₁</p>	<p>P₂ C₂</p>		<p>Medio término</p>
<p>4. <i>Comparación pre-prueba / post-prueba del grupo del proyecto, combinado con comparación post prueba de grupos del proyecto y de control.</i></p>	<p>P₁</p>	<p>X</p>		<p>P₂ C₁</p>		<p>Inicio</p>

						o
5. <i>Comparación post-prueba de grupos del proyecto y de control.</i> No se levantaron datos de base.		X		P ₁ C ₁		Fina l
DOS DISEÑOS DE EVALUACIÓN DÉBILES (NO-EXPERIMENTALES) PERO AMPLIAMENTE UTILIZADOS						
6. <i>Comparación pre-prueba / post-prueba de grupo del proyecto</i>	P ₁	X		P ₂		Inici o
7. <i>Análisis post-prueba de grupo del proyecto.</i> No se levantó información de base directa ni datos sobre un grupo de comparación.		X		P ₁		Fina l

Tabla 4 Fortalezas y debilidades de los siete diseños cuasi-experimentales más frecuentemente utilizados		
Diseño	Ventajas	Desventajas
1. Diseño longitudinal comprensivo con observaciones pre-prueba, de medio término, post-prueba y ex-post de grupos del proyecto y de comparación	Se trata del diseño más robusto, ya que estudia tanto el proceso de ejecución como la sostenibilidad del proyecto. Puede utilizarse para validar/estudiar una innovación introducida en un proyecto que, de probarse el impacto, se aplicará en mucho mayor escala.	La desventaja es que se trata del diseño más caro, oneroso y difícil de implementar.
2. Grupos del proyecto y de comparación pre-prueba y post-prueba	Se trata del diseño cuasi-experimental más robusto para todo propósito. Con un grupo de control adecuadamente seleccionado, genera buenos estimados de los impactos del proyecto.	<ul style="list-style-type: none"> ▪ Se asume que el grupo de comparación es razonablemente similar al grupo del proyecto y que está dispuesto a participar en dos encuestas pese a no recibir ningún beneficio a cambio. ▪ No se evalúa la ejecución del proyecto.
3. Diseño longitudinal truncado con grupos del proyecto y de comparación pre-prueba / post-prueba	<ul style="list-style-type: none"> ▪ Se observa el proceso de ejecución así como los impactos del proyecto. ▪ Modelo razonablemente robusto, especialmente para proyectos que empiezan lentamente, de modo tal que no se pierde mucho si la evaluación empieza tarde. 	<ul style="list-style-type: none"> ▪ No se empieza sino hasta cerca del medio término del proyecto y por lo tanto no se estudia el período de puesta en marcha ni la ejecución inicial del proyecto.
4. Grupo del proyecto pre-prueba / post-prueba combinado con análisis post-prueba de grupos del proyecto y de comparación	<ul style="list-style-type: none"> ▪ Sirve para evaluar si el enfoque del proyecto funciona y genera los productos deseados. ▪ Sirve para evaluar las similitudes y diferencias entre las áreas del proyecto y de control. ▪ Sirve para evaluar el potencial de réplica del proyecto. 	<ul style="list-style-type: none"> ▪ No se evalúa si las diferencias observadas al final del proyecto entre los grupos del proyecto y de comparación se deben al proyecto o si se trata de diferencias preexistentes entre ambos grupos. ▪ No se controla la historia local, que podría afectar los efectos directos.
5. Grupos del proyecto y de comparación post-prueba	<ul style="list-style-type: none"> ▪ Sirve para evaluar proyectos que promueven intervenciones bastante validadas u operan en áreas aisladas, donde no padecen interferencias por parte de otras intervenciones externas. 	<ul style="list-style-type: none"> ▪ No se estima la magnitud exacta de los impactos del proyecto. ▪ No se controla la historia local. ▪ No se evalúa el potencial de réplica a mayor escala. ▪ No se incide en el proceso de ejecución del proyecto.

<p>6. Grupo del proyecto pre-prueba / post-prueba</p>	<ul style="list-style-type: none"> ▪ Ofrece un estimado aproximado de los impactos del proyecto, especialmente en comunidades pequeñas o aisladas. 	<ul style="list-style-type: none"> ▪ No se estima la magnitud exacta de los impactos del proyecto. ▪ No se controla la historia local. ▪ No se compara el proyecto con otras comunidades. ▪ No se controla el efecto de las variables de intervención mediante el uso de un análisis multivariado.
<p>7. Grupo del proyecto post-prueba únicamente</p>	<ul style="list-style-type: none"> ▪ Útil para estudios exploratorios, para obtener una idea general de cuán bien funciona el modelo del proyecto. ▪ Ofrece un primer estimado aproximado de los resultados, especialmente en el caso de proyectos pequeños o aislados. 	<ul style="list-style-type: none"> ▪ No se obtienen estimados razonablemente precisos del impacto del proyecto. ▪ Es difícil tener la certeza de que los cambios observados se deben al proyecto, y no a otros factores o intervenciones. ▪ No se controlan los eventos externos. ▪ No se obtienen datos comparativos para estimar el potencial de réplica.

Nota: Se puede fortalecer la robustez de estos modelos combinándolos con el marco de evaluación de impacto y el análisis de factores contextuales que se presentan en el Capítulo 9), así como con algunas de las técnicas de EMR que tratamos en el Capítulo 10. En el caso de los diseños 1, 2, 3, 4 y 5, que utilizan grupos de comparación, se puede fortalecer el análisis sustancialmente empleando regresión múltiple para controlar estadísticamente las diferencias entre las características de los grupos del proyecto y de control. Ahí donde se disponga de datos secundarios, estos diseños pueden ser reforzados además mediante técnicas de apareamiento estadística, por ejemplo, apareamiento de puntajes de propensión y variables instrumentales (ver Banco Mundial, 2006).

Fuente: Bamberger, Rugh y Mabry, 2006. *RealWorld Evaluation*, Tabla 10.3.

Tabla 5. Reducción de los costos de recolección y análisis de datos para evaluaciones cuantitativas y cualitativas	
Evaluaciones cuantitativas	Evaluaciones cualitativas
A. Simplificar el diseño de evaluación	
<ul style="list-style-type: none"> • Comparación pre-prueba / post-prueba del grupo del proyecto con comparación post-prueba de grupos del proyecto y de control (Diseño 3): elimina el grupo de control de base. • Comparación pre-prueba / post-prueba de grupo del proyecto (Diseño 6): elimina el grupo de control pre-prueba y post-prueba. • Comparación post-prueba del grupo del proyecto y de control (Diseño 5): elimina la información de base. • Evaluación basada en datos post-prueba del grupo del proyecto (Diseño 7): elimina el grupo de control y el grupo de base del proyecto. 	<ul style="list-style-type: none"> • Priorizar los aspectos esenciales y enfocarse en ellos. • Reducir el número de visitas a los sitios del proyecto, o bien el período de tiempo durante el cual se realizan las observaciones. • Reducir la cantidad de datos recolectados y el costo de recolección de datos. • Reducir el número de personas o grupos estudiados.
B. Establecer las necesidades de información del cliente	
Priorizar las preguntas y las necesidades de datos junto con el cliente, con miras a eliminar la recolección de datos no realmente necesarios para los fines de la evaluación.	
C. Utilizar datos secundarias	
<ul style="list-style-type: none"> • Censos o encuestas realizados en las áreas del proyecto • Información obtenida de los registros del proyecto • Registros de escuelas, centros de salud y otras instituciones de servicio público 	<ul style="list-style-type: none"> • Diarios y otros medios de comunicación masiva • Registros de organizaciones comunitarias • Tesis de grado y postgrado y otras investigaciones universitarias [para evaluaciones CUALI y CUANTI]

D. Reducir el tamaño de la muestra	
<ul style="list-style-type: none"> • Reducir el nivel de precisión requerido (menor precisión = muestra mas pequeña) • Reducir los tipos de desglose requeridos (menos desglose = muestra más pequeña) • Diseños muestrales estratificados (menos entrevistas) • Utilizar muestreos por conglomerados (menos gasto por desplazamientos) 	<ul style="list-style-type: none"> • Considerar muestreos críticos o basados en cuotas, antes que muestreos comprensivos o representativos • Reducir el número de personas o grupos estudiados

E. Reducir los costos de recolección, ingreso y análisis de datos	
<ul style="list-style-type: none"> • Cuestionarios autoadministrados (en el caso de poblaciones alfabetizadas) • Observación directa (en vez de encuestas) (en ocasiones representa un ahorro de dinero, pero no en todos los casos) • Contadores automáticos y otros métodos no muy conspicuos • Ingreso directo de los datos de la encuesta en instrumentos portátiles • Escaneo óptico de los formularios de la encuesta y encuestas electrónicas 	<ul style="list-style-type: none"> • Reducir el número o el período de las observaciones • Priorizar a los informantes • Emplear y capacitar estudiantes universitarios, estudiantes de enfermería y pobladores de la comunidad para la recolección de datos (en evaluaciones tanto CUALI como CUANTI) • Ingreso de datos por medio de instrumentos portátiles
Diseños de métodos mixtos	
<ul style="list-style-type: none"> • Triangulación para compensar la reducción del tamaño de la muestra • Entrevistas con grupos focales y foros comunitarios en vez de encuestas de hogares • Diagnóstico Rural Participativo (DRP) y otros métodos participativos 	

Fuente: Bamberger, Rugh y Mabry, 2006. Tabla 3.1.

Tabla 6 AHORRO DE COSTOS ESTIMADO EN EL CASO DE EMR MENOS ROBUSTAS EN COMPARACIÓN CON EL DISEÑO 2

<i>Diseño</i>		<i>Ahorro de costos estimado en comparación con el Diseño 2</i>
3	Diseño longitudinal truncado	5–10%
4	Ausencia de información de base del grupo de comparación	10–20%
5	Ausencia de información de base de ambos grupos	30–40%
6	Ausencia de grupo de comparación	40–50%
7	Solo grupo del proyecto post-prueba	60–80%

NOTA: El ahorro de costos estimado se basa en la reducción porcentual del número total de entrevistas pero tomando en cuenta que hay costos fijos, por ejemplo el diseño del cuestionario y los costos de capacitación.

Tabla 7 Factores que afectan el tamaño de la muestra		
Factor	Explicación	Influencia en el tamaño de la muestra
1. El propósito de la evaluación	¿Se trata de un estudio exploratorio o se requieren estimados estadísticos muy precisos?	Cuanto más precisos sean los resultados requeridos, mayor será el tamaño de la muestra
2. ¿Se empleará una prueba unilateral o bilateral? (¿Se conoce la dirección del cambio esperado?)	Si el propósito de la evaluación es probar si los efectos directos positivos han aumentado, o si los efectos directos negativos se han reducido, en ese caso puede utilizarse una prueba unilateral. Si el propósito es probar si se ha producido “un cambio significativo” sin tener conocimiento de la dirección del mismo, en ese caso se requiere una prueba bilateral.	El tamaño de la muestra será aproximadamente 40 por ciento superior en el caso de una prueba bilateral.
3. ¿Se entrevistará solamente al grupo del proyecto?	En algunos diseños de evaluación solamente se entrevista a los sujetos del grupo del proyecto. Esto ocurre si se dispone de información sobre la población total basada en estudios anteriores o en datos secundarios. En otros casos se debe seleccionar y entrevistar además a un grupo de comparación.	Si es necesario entrevistar al mismo número de personas tanto en el grupo del proyecto como en el grupo de comparación se duplicará el tamaño de la muestra.
4. La homogeneidad del grupo	Si existe muy poca variación entre la población con respecto a la variable del efecto directo, en ese caso la desviación estándar será pequeña.	Cuanto más pequeña sea la desviación estándar, tanto más pequeña será la muestra.
5. El tamaño del efecto	El tamaño del efecto es el monto de incremento que se espera que produzca el proyecto.	Cuanto más pequeño sea el tamaño del efecto, tanto mayor será la muestra.
6. La eficiencia con la que se ejecuta el proyecto	Mientras que algunos proyectos son ejecutados de manera muy eficiente, en que todos los sujetos reciben exactamente el mismo paquete de servicios, en otros casos la administración es más precaria y distintos sujetos reciben distintas combinaciones de servicios. La calidad de los servicios también puede variar.	Cuanto peores sean la calidad y la eficiencia de la implementación del proyecto, tanto mayor será la muestra.
7. El nivel de desglose requerido	En algunos casos el cliente necesita solo estimados globales del impacto en la población total del proyecto. En otros se esperan resultados desglosados por sitios del proyecto, variaciones en el paquete de servicios provisto o grupos	Cuanto mayor sea el desglose requerido, tanto mayor será la muestra.

	socioeconómicos (sexo, edad, etnicidad, etc.).	
8. El diseño de la muestra	Los procedimientos muestrales, por ejemplo el método de estratificación, a menudo pueden reducir la varianza de los estimados y aumentar la precisión.	Una estratificación bien diseñada puede reducir el tamaño de la muestra.
9. El nivel de precisión estadística	Cuando se estima si un proyecto ha tenido un impacto “más allá de toda duda razonable”, este concepto suele definirse como “una posibilidad menor a una en 20 de que un impacto tan grande como éste pudiera haber ocurrido por casualidad”. Esto se define al nivel de confianza de 0.05. Si se requieren resultados más precisos se puede utilizar un nivel de confianza de 0.01 (menos de uno en 100). En estudios exploratorios se puede utilizar el nivel de confianza de 0.10 (uno en 10).	Cuanto más alto sea el nivel de confianza, tanto mayor será la muestra.
10. El poder de la prueba	El poder estadístico de la prueba se refiere a la probabilidad de que, cuando un proyecto tiene un efecto “real”, éste sea rechazado por la prueba de significación estadística. El nivel de poder convencional es de 0.8, lo cual quiere decir que apenas hay un 20% de posibilidades de que un efecto real sea rechazado. Ahí donde se requiere un nivel de precisión más alto, el poder puede incrementarse a 0.9 o más.	Cuanto mayor sea el nivel de poder, tanto mayor será la muestra.
11. Factor de corrección para población finita	Si la muestra representa un porcentaje mayor a, digamos, el cinco por ciento de la población total, es posible reducir el tamaño de la muestra mediante el factor de corrección para población finita.	Cuanto mayor sea el porcentaje de la población total que representa la muestra, menor será la muestra.

Fuente: Bamberger, Rugh y Mabry, 2006. Tabla 3.3.

Tabla 8. Reducción del tiempo requerido para la recolección y el análisis de datos en evaluaciones cuantitativas y cualitativas

Enfoques también utilizados para reducir los costos		
1. Simplificar el diseño de evaluación		
2. Establecer y priorizar las necesidades de información del cliente		
3. Utilizar información documentaria ya existente		
4. Reducir el tamaño de la muestra		
5. Utilizar métodos de recolección de datos más rápidos y económicos		
Enfoques adicionales que representan un ahorro de tiempo pero no de dinero y a menudo incrementan los costos		
	Cuantitativas	Cualitativas
<p>6. Reducir presiones sobre el de tiempo de los consultores/as o subcontratistas externos (a menudo extranjeros)</p> <p>a. Encargar a consultores/as locales la recolección y organización anticipada de los datos disponibles.</p> <p>b. Encargar estudios exploratorios a un consultor/a local, quien deberá identificar algunos de los problemas clave así como las características de la población antes de la llegada del consultor/a externo.</p> <p>c. Por medio de videoconferencias entre el consultor/a externo y el consultor/a local antes de la visita del consultor/a externo, se puede avanzar con la planificación con el objeto de ahorrar tiempo.</p>	<p>a. Recopilación de datos secundarios y evaluación inicial de la calidad y la pertinencia de los mismos para el estudio (Nota: el mismo tipo de documentos ya existentes se emplearía en investigaciones CUALI.)</p> <p>b. <u>Encuestas rápidas</u> para obtener datos demográficos, económicos y otros datos pertinentes sobre las poblaciones objetivo, para ayudar a preparar el diseño de la muestra y el marco muestral (una lista o mapa con la ubicación geográfica de todas las familias u otros sujetos de la población estudiada). Los estudios rápidos pueden utilizarse además para obtener estimados preliminares de, por ejemplo, puntajes de educación o alfabetismo.</p> <p>c. Establecer una buena relación con la comunidad y con líderes y funcionarios/as locales para facilitar la</p>	<p>a. Recopilación de material de consulta y fuentes de investigación tales como materiales de medios de comunicación masiva, fotografías, etc.</p> <p>b. <u>Estudios etnográficos rápidos</u>, concentrándose en los conceptos y temas clave a ser abordados por el estudio y como trabajo preparatorio para los consultores/as externos.</p> <p>c. Fotos, videos y grabaciones que puedan ser enviadas a los consultores/as externos para documentar la situación de las comunidades en distintas épocas del año (por ejemplo, durante los monzones y en la temporada seca). Esto puede ser importante si a los consultores/as no les es posible visitar la región en diversas temporadas.</p> <p>c. Establecer una buena relación (igual que en el caso de evaluaciones CUANTI).</p>

	ejecución sin contratiempos del estudio y evitar demoras burocráticas (por ejemplo, conseguir los documentos requeridos para iniciar el estudio).	
<p>7. Contratar un número mayor de recolectores/as de datos.</p> <p>a. Aumentar el número de entrevistadores/as y supervisores/as.</p> <p>b. Contratar entrevistadores/as y supervisores/as más experimentados. Esto puede reducir el tiempo invertido en capacitación y aumentar la eficiencia y la velocidad de los recolectores/as regulares de datos.</p> <p>c. Subcontratar la recolección o el análisis de los datos.</p>		
<p>8. Revisar el formato de los registros del proyecto con el fin de incluir información crucial para el análisis del impacto.</p>	Incluir indicadores sobre el acceso a, y el uso de, servicios, así como características pertinentes ya sea a nivel individual o de hogares – especialmente indicadores de impacto.	Incentivar/permitir al personal del proyecto registrar algo más que las actividades del proyecto, si fuese pertinente para la evaluación, para documentar la observación de cambios en la situación de los hogares de la población beneficiaria.
<p>9. Tecnología moderna de recolección y análisis de datos.</p>	<ul style="list-style-type: none"> • Computadoras portátiles para el ingreso de los datos • Escaneo óptico • Contadores automáticos • Programas de computación CUANTI para el análisis y la presentación de los datos (Nota: establecer si esto realmente significa un ahorro de tiempo, ya que el propósito principal es hacer posible un análisis 	<ul style="list-style-type: none"> • Encuestas por correo electrónico (se pueden usar en estudios tanto CUALI como CUANTI) • Videocámaras y grabadoras • Fotografía • Se pueden utilizar mapeos GPS y fotografías aéreas para observar los patrones demográficos, las prácticas agropecuarias y la situación de la infraestructura a lo largo de un período

	más comprensivo)	de tiempo extendido, lo cual puede representar un ahorro frente a la considerable cantidad de tiempo que se invierte en llegar a poblados y zonas distantes <ul style="list-style-type: none">• Programas de computación CUALI para el análisis y la presentación de los datos
--	------------------	--

Tabla 9 Métodos de recolección rápida de datos

		Ahorro de tiempo transcurrido, esfuerzo o ambos		
Opciones para reducir los requerimientos de tiempo		Ti empo transcurrido	E sfuerzo	A mbos
A. Métodos eminentemente cualitativos				
Entrevistas con informantes clave	<p>Los informantes clave pueden representar un ahorro de tiempo, ya sea proporcionando información (precios de productos agropecuarios, personas que abandonan o ingresan a la comunidad, asistencia y ausentismo escolar), ayudando a los investigadores/as a enfocarse en los temas clave o identificando formas más rápidas de obtener información. Opciones para reducir el tiempo de las entrevistas con informantes clave:</p> <ul style="list-style-type: none"> • Reducir el número de informantes. • Limitar el número de temas cubiertos. • Contratar un número mayor de investigadores/as para llevar a cabo las entrevistas o grabar las entrevistas para que sean visualizadas por el investigador/a. Hacer esto con cautela, recordando que es importante para el investigador/a mantener un contacto personal con personas clave de la comunidad. 			
Entrevistas con grupos focales y con la comunidad	<ul style="list-style-type: none"> • Subcontratar especialistas en entrevistas con grupos focales, por ejemplo, una empresa de investigación de mercado. • Llevar a cabo varias entrevistas con grupos focales simultánea en vez de consecutivamente. • Recolectar información de reuniones antes que de encuestas. Se puede recabar información sobre temas tales como el acceso y el uso de abastecimiento de agua y el saneamiento, las prácticas agropecuarias y la división de género del trabajo en la agricultura a través de entrevistas grupales, posiblemente combinadas con la distribución de encuestas autoadministradas. 	√		

	<ul style="list-style-type: none"> • Es importante emplear técnicas para asegurar que se recojan los puntos de vista de todos los y las participantes (presiones de tiempo pueden llevar a que se deje de lado a los grupos más vulnerables y difíciles de acceder). 			
Observación estructurada	<ul style="list-style-type: none"> • En ocasiones, pero no siempre, la observación puede ser más rápida que las encuestas. Por ejemplo: observación de la división de género del trabajo en distintos tipos de actividades agropecuarias, quiénes asisten a las reuniones y participan en las discusiones, tipos de conflictos observados en los espacios públicos de la comunidad. 	√		
Uso de documentos y artefactos preexistentes	<ul style="list-style-type: none"> • Muchos tipos de datos preexistentes pueden ser más rápidamente recolectados y revisados que datos nuevos. Por ejemplo: registros de asistencia escolar, periódicos y otros medios de comunicación masiva, actas de asambleas comunales, registros de centros de salud, encuestas en comunidades objetivo realizadas por institutos de investigación. 	√		
Uso de grupos comunitarios para recolectar información	<ul style="list-style-type: none"> • Organización de estudios rápidos sobre la comunidad (CUALI y CUANTI) utilizando entrevistadores/as de la comunidad (los maestros/as de escuelas locales a menudo colaboran con estas actividades) 			
Fotos y videos	<ul style="list-style-type: none"> • Distribuir cámaras o videocámaras desechables a informantes de la comunidad para que tomen fotos (o graben videos) ilustrando, por ejemplo, los problemas de la comunidad. 	√		
Triangulación	<ul style="list-style-type: none"> • Disponer que varios entrevistadores/as realicen entrevistas simultáneamente y registren por separado sus observaciones sobre los mismos entrevistados/as clave, en vez entrevistar por separado. Esto puede representar un ahorro de tiempo transcurrido si sustituye la ejecución de varias entrevistas separadas con la misma persona. 	√		
B. Métodos eminentemente cuantitativos				
Encuestas rápidas basadas en cuestionarios breves y muestras pequeñas	<ul style="list-style-type: none"> • Reducir el número de preguntas y el tamaño de la muestra puede reducir significativamente el tiempo requerido para llevar a cabo una encuesta. • Incrementar el número de entrevistadores/as. 			
Reducir el tamaño de las muestras	<ul style="list-style-type: none"> • Existen técnicas muestrales especializadas, tales como el muestreo de aceptación por lotes (Valadez y Bamberger, 1994), diseñadas para generar estimados del uso o la calidad de servicios públicos como salud y educación con base en muestras muy pequeñas. Muestras de 14 a 28 hogares pueden ser suficientes para evaluar el uso o la calidad de un centro de salud. 			

Triangulación (utilizada también en métodos CUALI y mixtos)	<ul style="list-style-type: none"> • Obtener estimados independientes de distintas fuentes (por ejemplo, encuesta y observación) en ocasiones permite obtener estimados de muestras más pequeñas, ahorrando con ello en términos tanto de tiempo transcurrido como de esfuerzo. 			
Encuestas rápidas de salida	<ul style="list-style-type: none"> • Se pregunta a las personas que están saliendo de una reunión o de un local de servicios que escriban sus puntos de vista sobre la reunión o el servicio en una ficha colgada de la pared. • A menudo se hace una sola pregunta. Por ejemplo: “¿Le recomendaría a un vecino o vecina que acuda a la próxima reunión o que utilice este centro?” 			
Uso de datos preexistentes	<ul style="list-style-type: none"> • Las encuestas previas u otras fuentes de datos pueden tornar innecesaria la recolección de ciertos datos. • Los hallazgos de encuestas previas pueden reducir el tiempo requerido para diseñar una muestra o, al proporcionar información sobre la <i>desviación estándar</i> (cuán estrecha o ampliamente distribuidos se encuentran los sujetos alrededor de la media) de variables clave, pueden facilitar la reducción del tamaño de la muestra o el ahorro de tiempo si se utilizan <i>muestreos por estratificación o por conglomerados</i> más eficientes. (En el Capítulo 16 se definen estos términos). 	√		
Listas de verificación para observaciones	<ul style="list-style-type: none"> • Las listas de verificación para observaciones con frecuencia eliminan la necesidad de ciertas encuestas (por ejemplo, flujos del tráfico pedestre o vehicular, uso de instalaciones y servicios comunitarios, tiempo requerido para recolectar agua y combustible). 			
Contadores automáticos	<ul style="list-style-type: none"> • Registran a las personas que ingresan a un local o utilizan un servicio (por ejemplo, una instalación de agua). 			
C. Métodos mixtos				
Triangulación (utilizada también en métodos CUALI y CUANTI)	<ul style="list-style-type: none"> • Triangular los datos obtenidos mediante una combinación de métodos cuantitativos y cualitativos en ocasiones permite obtener estimados basados en muestras más pequeñas, ahorrando con ello esfuerzo y tiempo transcurrido. Nota: no siempre es el caso, dado que la aplicación de un número mayor de métodos de recolección de datos tiene implicaciones obvias en términos de tiempo/costo. 			
Cuantificación rápida de <i>métodos de evaluación</i>	<ul style="list-style-type: none"> • Se pueden combinar encuestas breves y rápidas con estimados numéricos obtenidos de entrevistas con la comunidad y grupos focales para generar estimados de, por ejemplo, el uso 			

<p><i>participativa</i> y entrevistas con grupos focales</p>	<p>de los servicios, las tasas de desempleo, el <i>uso del tiempo</i> en una comunidad u otro grupo poblacional.</p>			
<p>Nota: A menudo resulta difícil diferenciar entre ahorro de tiempo y reducción del esfuerzo requerido. Además, es importante tener presente que ahorrar tiempo incrementando el tamaño del equipo generalmente eleva el presupuesto. De ahí la necesidad de establecer junto con el cliente si la restricción principal es el tiempo, el presupuesto o ambos.</p>				

Tabla 10 Estrategias para responder a las restricciones de información

Reconstrucción de los datos de base^a		
<i>Enfoques</i>	<i>Fuentes/Métodos</i>	<i>Comentarios/Problemas</i>
Utilizar documentos existentes (datos secundarios)	<ul style="list-style-type: none"> • Registros del proyecto • Datos de instituciones de servicio público (salud, educación, etc.) • Encuestas de hogares y otras encuestas relacionadas 	
Evaluar la confiabilidad y la validez de los datos secundarios (ver el Capítulo 8 para una discusión de estos conceptos)	<ul style="list-style-type: none"> • Registros de matrícula y asistencia escolares • Historias médicas de pacientes en centros de salud locales • Registros de préstamos y devoluciones en cooperativas de ahorro y crédito • Registros vehiculares (para estimar los cambios en el volumen del tráfico) • Registros de mercados de agricultores locales (precios y volumen de ventas) 	<p>Todos los datos deben ser evaluados para determinar su idoneidad en términos de</p> <ul style="list-style-type: none"> • Período de referencia • Cobertura demográfica • Inclusión de indicadores requeridos • Documentación sobre las metodologías utilizadas • Si están completos • Exactitud • Si están libres de sesgos
La retrospectiva: solicitar a las personas que suministren datos numéricos (ingreso, producción de cultivos, cuántas horas al día dedican a desplazarse, tarifas escolares) o cualitativos (nivel de violencia en la comunidad, nivel de consulta con la comunidad por parte de los funcionarios/as del gobierno local) sobre cómo era la situación alrededor de la fecha de inicio del proyecto	<ul style="list-style-type: none"> • Informantes clave • DRP (diagnóstico rural participativo) y otros métodos participativos 	<p>La retrospectiva puede emplearse para recolectar datos sobre</p> <ul style="list-style-type: none"> • Asistencia escolar • Enfermedades/uso de servicios e instalaciones de salud • Ingreso/ganancias • Conocimiento y destrezas a nivel comunitario/individual • Cohesión social y conflictos • Consumo y costo del agua • Gastos mayores o rutinarios a nivel de hogares • Períodos difíciles • Patrones de desplazamiento y transporte de productos agropecuarios
Mejorar la confiabilidad y/o la	<ul style="list-style-type: none"> • Efectuar pequeños estudios pre-prueba / post- 	

<p>validez de los datos recordados</p>	<p>prueba para cotejar los datos recordados con información original</p> <ul style="list-style-type: none"> • Identificar y tratar de controlar potenciales sesgos (subestimación de gastos pequeños, distorsión de gastos mayores al incluir algunos gastos realizados con anterioridad al período de recordación, distorsión de la información para encajar con la norma de conducta aceptada, intención de engañar al entrevistador/a o llevarlo a conclusiones equivocadas) • Explicar el contexto (el período de tiempo, ciertos tipos de conductas específicos, las razones para recolectar la información) • Asociar la retrospectiva acerca de eventos referenciales importantes en la historia de la comunidad o persona • Triangulación (informantes clave - fuentes secundarias - DRP) 	
<p>Informantes clave</p>	<ul style="list-style-type: none"> • Dirigentes comunitarios • Líderes religiosos • Maestros/as • Doctores/as y enfermeros/as • Propietarios de tiendas • Policías • Periodistas 	<ul style="list-style-type: none"> • Utilizarlos para triangular (validar la consistencia de) los datos obtenidos de otras fuentes
<p>Problemas y desafíos especiales cuando se trabaja con grupos de comparación</p>		
<p><i>Enfoque</i></p>	<p><i>Fuentes</i></p>	<p><i>Comentarios/Problemas</i></p>
<p>Identificar y reconstruir grupos de comparación</p>	<p>Gobierno, estadísticas, encuestas anteriores, registros escolares, centros de salud y otras instituciones de servicio público</p>	<p>Los desafíos y problemas incluyen:</p> <ul style="list-style-type: none"> • Presiones políticas • Consideraciones éticas relacionadas con el uso de grupos de comparación • Uso de encuestas anteriores como marco muestral • Estudios piloto rápidos

		<p>para probar la varianza, etc.</p> <ul style="list-style-type: none"> • Correspondencia basada en el juicio del evaluador/a • Uso de fases posteriores del proyecto con fines de comparación • Grupos de comparación internos, cuando distintos participantes reciben distintas combinaciones de servicios • Idoneidad de potenciales grupos de comparación • Creación estadística de control (análisis de conglomerados, apareamiento por puntajes de propensidad)
<p>Problemas especiales con la reconstrucción de datos sobre los grupos de comparación</p>	<ul style="list-style-type: none"> • La comparación econométrica post-prueba de las áreas del proyecto y de comparación no permite controlar las diferencias históricas entre los dos grupos (ver el Capítulo 10). 	<p>Problemas metodológicos:</p> <ul style="list-style-type: none"> • Autoselección de los participantes (problemas: dificultad de establecer una correspondencia con un grupo de comparación respecto de factores tales como motivación) • Proyectos seleccionados para representar ya sea a los grupos con el mayor potencial de éxito o los grupos que enfrentan el mayor desafío (problemas: en ambos casos, dificultad de encontrar un grupo de comparación con características similares)
<p>Recolección de datos sensibles (por ejemplo, violencia doméstica, conductas sobre fertilidad, toma de decisiones y control de los recursos en el hogar, información de o sobre mujeres, así como información sobre personas con discapacidad física o mental)</p>	<ul style="list-style-type: none"> • Observación participante • Grupos focales • Entrevistas no estructuradas • Observación • Técnicas de DRP • Estudios de caso • Informantes clave 	<p>Estos problemas se dan también con los participantes del proyecto, pero tienden a ser más difíciles de resolver en el caso de los grupos de comparación porque el investigador/a no tiene los mismos contactos o el mismo acceso a la comunidad.</p>
<p>Recolección de datos</p>	<ul style="list-style-type: none"> • Observación (participante) 	

<p>sobre grupos difíciles de acceder (por ejemplo, trabajadores/as en el comercio sexual, consumidores de drogas o alcohol, delincuentes, pequeños negocios informales, invasores y residentes ilegales, minorías étnicas o religiosas, así como, en algunas culturas, las mujeres)</p>	<p>y no participante)</p> <ul style="list-style-type: none"> • Informantes de los grupos • Autoinformación • Estudios de seguimiento (<i>tracer studies</i>) y muestreos “bola de nieve” • Informantes clave • Documentos existentes (datos secundarios) • Símbolos de identificación grupal (ropa, tatuajes, graffiti) 	<p>Igual que en el punto anterior.</p>
---	---	--

Tabla 11 Factores que determinan la idoneidad del diseño y de los hallazgos de la evaluación

<p>1. Qué tan adecuados son la focalización, el enfoque y los métodos de la evaluación para obtener la información requerida sobre, por ejemplo:</p> <ul style="list-style-type: none"> a. Decisiones gerenciales b. Opinión de los interesados sobre la idoneidad del programa
<p>2. Cuán disponibles se encuentran los datos y las fuentes de datos, por ejemplo:</p> <ul style="list-style-type: none"> a. Si existen o pueden generarse datos apropiados para satisfacer las necesidades de información. b. Si los interesados y las fuentes de datos documentarios son accesibles para los evaluadores/as.
<p>3. Cuán bien respaldarán los datos interpretaciones válidas sobre el programa en relación con, por ejemplo:</p> <ul style="list-style-type: none"> a. El logro de las metas del programa, el grado de obtención de los beneficios del programa b. El costo-efectividad del programa d. La idoneidad de los recursos que afectan el logro de las metas e. Las consecuencias no intencionales del programa
<p>4. Cuán adecuado es el equipo de evaluación, por ejemplo en términos de:</p> <ul style="list-style-type: none"> a. La metodología de evaluación b. El campo específico del programa c. Si los recursos de evaluación son suficientes para las dimensiones del programa

Tabla 12 Algunas amenazas que afectan la validez incluso de los diseños cuantitativos más robustos (Diseños 1 y 2)

<i>Amenazas a la validez^a</i>	
F. Amenazas a la validez de la conclusión estadística	
1. <i>Bajo poder estadístico</i>	La muestra es muy pequeña como para poder detectar efectos estadísticamente significativos (ver Capítulo 14).
4. <i>Poca confiabilidad de las medidas</i>	Los indicadores no miden adecuadamente variables clave.
5. <i>Restricción del rango</i>	La muestra no cubre a toda la población. Por ejemplo, los grupos de menor o mayor ingreso se encuentran excluidos de la muestra, o ésta solo cubre a las empresas que emplean a más de 10 personas.
6. <i>Poca confiabilidad de la aplicación del tratamiento</i>	Los tratamientos no fueron aplicados uniformemente a todos los sujetos, y a menudo las diferencias en la aplicación de los mismos no están documentadas. Por ejemplo, algunas madres recibieron pastillas contra la malaria y orientación de parte del enfermero o enfermera, mientras que otras solo recibieron las pastillas.
G. Amenazas a la validez interna	
2. <i>Sesgos en la selección</i>	Diferencias entre los grupos del proyecto y de comparación relacionadas con factores que afectan los efectos directos.
6. <i>Atrición</i>	Aunque el grupo del proyecto inicialmente haya sido representativo de la población total, algunos subgrupos (por ejemplo, los menos educados, las mujeres con hijos pequeños, los trabajadores/as independientes) acusan tasas más altas de deserción y por lo tanto las personas que realmente han estado expuestas al proyecto dejan de ser representativas de toda la población.
H. Amenazas a la validez del marco conceptual	
1. <i>Explicación inadecuada de los marcos conceptuales y del modelo teórico del programa</i>	No se explican o definen claramente los conceptos básicos del modelo.
8. <i>Reactividad ante los instrumentos de recolección de datos</i>	Las respuestas pueden verse afectadas por la forma en que los sujetos reaccionan ante la entrevista u otros métodos de recolección de datos. Por ejemplo, los entrevistados pueden informar que son más pobres de lo que realmente son, o que el proyecto no ha generado beneficios porque esperan que la organización provea nuevos servicios o reduzca el costo de los servicios actuales.
I. Amenazas a la validez externa	
6. <i>Influencia de quienes diseñan las políticas en los efectos directos del programa</i>	El respaldo o la oposición de quienes diseñan las políticas en determinadas localidades pueden afectar los efectos directos del programa de formas que pueden ser difíciles de evaluar.
7. <i>Ciclos estacionales</i>	Muchas encuestas se llevan a cabo durante una sola época del año y pueden no recoger adecuadamente variaciones estacionales importantes.

a. Ver la “Lista de verificación integral para evaluar las amenazas a la validez de los diseños cuantitativos, cualitativos y de métodos mixtos” (Anexo 1) para una relación

completa de las amenazas. Los números que aparecen en la columna de la izquierda corresponden a dicha lista.

Tabla 13 Características de los enfoques CUANTI y CUALI en diferentes etapas del proceso de evaluación

<i>Actividad de evaluación</i>	<i>Enfoque cuantitativo</i>	<i>Enfoque cualitativo</i>
Marco conceptual y formulación de hipótesis	<ul style="list-style-type: none"> • Generalmente, pero no siempre, las evaluaciones se basan en un marco teórico que se construye a partir de una revisión de la bibliografía y suele comprender hipótesis comprobables. • Las hipótesis a menudo son <i>deductivas</i> (basadas en hipótesis comprobables derivadas de la teoría). • Las hipótesis suelen ser cuantitativas y pueden ser evaluadas mediante pruebas de significación estadística. • El marco con frecuencia empieza en el nivel macro, antes que micro. 	<ul style="list-style-type: none"> • Aunque algunas evaluaciones definen y validan hipótesis, muchas no lo hacen. • Muchas evaluaciones hacen hincapié en la singularidad de cada situación, y el marco conceptual se define mediante un proceso de iteración, en que el marco es constantemente actualizado a medida que se obtiene nueva información. • Las hipótesis, si se plantean, frecuentemente son <i>inductivas</i> (derivadas de la información recolectada en el transcurso de la investigación).
Selección de los sujetos o unidades de análisis	<ul style="list-style-type: none"> • Los muestreos aleatorios permiten generalizar los hallazgos y validar estadísticamente las diferencias entre los grupos. • Se necesita un marco muestral que contenga una relación de todos los miembros de la población o poblaciones objetivo a ser estudiadas. • Los métodos de selección suelen definirse con anticipación, se encuentran claramente documentados y se mantienen incólumes a lo largo de la investigación. • Típicamente se selecciona una muestra bastante grande de la cual se recolecta un conjunto finito de datos cuantitativos. 	<ul style="list-style-type: none"> • La decisión sobre los procedimientos de selección varía de acuerdo al propósito del estudio. • Se utiliza un muestreo intencionado para recolectar los datos más útiles e interesantes, relacionados con el propósito del estudio. • Aunque no suele hacerse en evaluaciones CUALI, en ocasiones en los enfoques de métodos mixtos se puede seleccionar la muestra empleando el mismo marco muestral maestro usado para el componente CUANTI de la investigación. Por ejemplo, para el análisis CUALI se puede seleccionar una submuestra de las aldeas de las cuales se seleccionaron las muestras de hogares (u otras unidades) para la encuesta CUANTI (aunque los tipos de recolección de datos y los sujetos, grupos u organizaciones a ser estudiados en el análisis CUALI tenderán a diferir). • Se suele entrevistar a un número más pequeño de personas pero en mayor profundidad.

<p>Diseño de evaluación</p>	<ul style="list-style-type: none"> • Normalmente se utiliza uno de los diseños cuasi-experimentales descritos en el Capítulo 10. Se entrevista a una muestra seleccionada de manera aleatoria que representa a los participantes del proyecto y, posiblemente, a un grupo de control o comparación en uno o más momentos durante la vida del proyecto. • Ahí donde sea posible se estiman los efectos directos (impactos) comparando los datos recolectados antes y después (y posiblemente durante) la ejecución del proyecto. 	<ul style="list-style-type: none"> • El investigador/a o investigadores/as se insertan en la comunidad durante un período de tiempo largo. • Se estudian los efectos del programa recolectando información sobre los numerosos elementos distintos de la comunidad así como sobre el contexto económico, político, cultural, ecológico y psicológico. • La evaluación normalmente no intenta establecer una relación causa-efecto o una relación lineal directas.
<p>Recolección de datos y métodos de registro</p>	<ul style="list-style-type: none"> • Los datos suelen registrarse en cuestionarios estructurados a los cuales se ciñen los encuestadores constantemente a todo lo largo de la investigación. • Existe un uso extendido de preguntas cerradas precodificadas. • El estudio utiliza principalmente valores numéricos (variables <i>integer</i>) o variables cerradas (ordinales o nominales) que pueden ser sometidas al análisis estadístico. • En las observaciones se pueden utilizar listas de verificación con respuestas precodificadas. 	<ul style="list-style-type: none"> • Los protocolos de las entrevistas son el instrumento más común, con frecuencia semiestructurado. • El instrumento de recolección de datos puede ser modificado en el transcurso de la investigación, a medida que se desarrolla un mayor conocimiento. • Los datos obtenidos en las entrevistas en ocasiones son respuestas literales grabadas (cintas de audio, cintas de video) y en ocasiones anotaciones escritas. • Se puede emplear en la investigación el análisis de documentos existentes (análisis documentario). Los datos textuales de los documentos a menudo se resaltan en una copia del original, que se adjunta como parte del conjunto de datos. • Se pueden emplear entrevistas con grupos focales en el estudio (generalmente de menos de 10 personas), así como reuniones con grupos relativamente grandes de la comunidad. • Se puede utilizar en la investigación la observación participante y no participante. • Se puede utilizar en la investigación la fotografía.
<p>Triangulación</p>	<ul style="list-style-type: none"> • Los cuestionarios cuentan con controles de consistencia incorporados a fin de arrojar estimados independientes de variables clave (por ejemplo, se pueden cotejar datos sobre el ingreso con datos sobre los gastos). • Se puede utilizar la observación directa (una técnica 	<ul style="list-style-type: none"> • Se emplean varios métodos cualitativos para obtener perspectivas múltiples y triangular la información. • Triangulación por observación: Un monitor/a puede observar a un grupo focal o una reunión grupal, tanto para identificar cualesquier potenciales sesgos que resulten de la

	<p>CUALI) como un control de la consistencia de las respuestas proporcionadas por los encuestados (por ejemplo, se puede cotejar la información sobre el ingreso con evidencias sobre el número y la calidad de bienes de consumo duraderos dentro y fuera de la vivienda).</p> <ul style="list-style-type: none"> • En ocasiones se utiliza información de encuestas anteriores con los mismos encuestados, como un control de la consistencia de la información obtenida en una encuesta posterior. • Se pueden utilizar datos secundarios (datos censuales, encuestas de hogares a nivel nacional, información de instituciones del gobierno) para corroborar los estimados de la encuesta de la evaluación. 	<p>forma en que se llevó a cabo la reunión como para ofrecer una perspectiva independiente (por ejemplo, informando sobre las interacciones entre los miembros del grupo, observando cómo determinadas personas responden a los comentarios o la conducta de otras, etc.).</p>
<p>Análisis de datos</p>	<p>Ver el Capítulo 11 para mayores detalles sobre este y otros métodos.</p> <ul style="list-style-type: none"> • Se puede utilizar en la investigación el análisis de las preguntas que no fueron respondidas a fin de determinar si existe algún sesgo sistemático (por ejemplo, tasas más altas de preguntas sin responder entre participantes de menores o mayores ingresos, tasas de respuesta distintas de parte de los hombres y las mujeres). • Se puede utilizar en la investigación el análisis de la información faltante y las variables con valores fuera de los rangos esperados; es necesario decidir si excluir o no dichos valores (con los consiguientes sesgos resultantes en las estimaciones) o realizar ajustes estadísticos. • Se pueden utilizar en la investigación estadísticas descriptivas – indicadores de la tendencia central y de la dispersión y la tendencia central (ver el Capítulo 11) • Se puede emplear en la investigación un análisis multivariado para examinar los factores que contribuyen a la magnitud y la dirección del cambio. • Se pueden emplear en la investigación pruebas de 	<p>Ver el Capítulo 12 para una discusión sobre la mayoría de estos métodos.</p> <ul style="list-style-type: none"> • Se puede utilizar en la investigación el análisis inductivo a partir de los datos recolectados, a fin de analizar patrones que permitan inferir deducciones y construir interpretaciones. • Se puede utilizar en la investigación el análisis del contenido temático de las entrevistas, informes de las observaciones, documentos y otras fuentes. • El análisis de casos únicos plantea que cada caso tiene atributos singulares y que se puede aprender mucho concentrando el análisis en un solo caso. • Se puede utilizar en la investigación el análisis de casos cruzados, que ayuda a evaluar la medida en la que los hallazgos de casos individuales son generalizables y a dilucidar los procesos y efectos directos de muchos casos. • Se puede utilizar en la investigación el método de comparación constante, según el cual se comparan constantemente datos nuevos con interpretaciones preliminares. • El análisis CUALI es holístico, y el programa que está

	<p>significación para determinar las diferencias entre distintos grupos.</p>	<p>siendo estudiado es considerado como un complejo tapiz de hilos entrelazados. El análisis enfatiza el contexto (entorno) y cómo éste afecta el funcionamiento del programa que está siendo evaluado.</p> <ul style="list-style-type: none">• El análisis CUALI implica una comprensión intuitiva.• Se puede trabajar con estudios de casos para estudiar las variables con valores fuera del rango esperado o ayudar a explicar y entender el programa en todos sus alcances.
--	--	---

Tabla 14 Elementos de un enfoque de investigación multidisciplinario integral

Composición del equipo de investigación

- Incluya investigadores/as primarios de distintas disciplinas. Programe un tiempo adecuado para que los investigadores/as se familiaricen con, y desarrollen respeto por, las disciplinas y el trabajo de los otros miembros del equipo. Cada investigador/a deberá familiarizarse con la bibliografía básica y los debates actuales en la disciplina de los demás.
- Cerciórese de que se establezcan vínculos como los arriba mencionados entre los investigadores/as locales de la ciudad, estado o región donde se está ejecutando el proyecto.

Enfoques integrales durante el diseño de la evaluación

- Cerciórese de que el marco de evaluación se base en teorías y enfoques de todos los equipos disciplinarios que participan en la evaluación (por ejemplo, antropología, medicina, derecho, sociología, economía, demografía), así como en marcos sustentados en perspectivas predominantemente cualitativas y cuantitativas, en que cada uno sea utilizado para enriquecer y fortalecer al otro.
- Cerciórese de que las hipótesis y los enfoques de investigación aprovechen elementos de todas las disciplinas.
- El marco de investigación debe formular los eslabonamientos entre los diferentes niveles de análisis (por ejemplo, tanto una encuesta cuantitativa como entrevistas cualitativas de hogares, estudiantes, agricultores; análisis cualitativo holístico del contexto del programa).
- Cerciórese de que los conceptos y métodos no sean aplicados fuera de contexto, sino que se sustenten, por el contrario, en los debates y enfoques intelectuales que están teniendo lugar al interior de las respectivas disciplinas.
- Considere utilizar modelos de conducta que combinen modelos económicos y otros modelos cuantitativos con una comprensión en profundidad del contexto cultural dentro del cual se está desarrollando la evaluación.

Recolección de datos y triangulación

- Realice un análisis exploratorio para ayudar a formular las hipótesis y definir los indicadores.
- Seleccione métodos de recolección de datos cuantitativos y cualitativos diseñados para complementarse mutuamente, y especifique las complementariedades entre ambos y cómo se emplearán en el trabajo de campo y en el análisis.
- Seleccione al menos dos métodos de estimación independientes para indicadores e hipótesis clave.
- Cerciórese de que todos los métodos de selección de muestras, recolección de datos y análisis se encuentren plenamente documentados.

Análisis e interpretación de datos y posible seguimiento ex post en el campo

- Efectúe y presente análisis separados de los hallazgos cuantitativos y cualitativos con el fin de resaltar diferentes interpretaciones y hallazgos obtenidos por diferentes métodos, y luego prepare un informe integral que aproveche todos los datos.
- Utilice sistemáticamente procedimientos de triangulación para verificar las inconsistencias o

interpretaciones divergentes. Realice un seguimiento ex post de las diferencias, ahí donde sea necesario, con una nueva visita al campo.

- Presupueste los recursos y programe los días necesarios para realizar visitas de seguimiento ex post al campo (no solo para métodos mixtos).
- Resalte las diferentes interpretaciones y los hallazgos obtenidos por diferentes métodos y provoque una discusión sobre cómo éstos enriquecen la interpretación del estudio. Los efectos diferentes deben ser considerados como una fortaleza importante de un enfoque integral y no como un fastidio.
- Presente casos y material cualitativo para ilustrar o validar los hallazgos cuantitativos.

Presentación y difusión de los hallazgos

- Combine formas de presentación convencionales, por ejemplo, informes escritos complementados por presentaciones en PowerPoint, con algunos de los métodos de presentación más participativos que se utilizan en algunas evaluaciones cualitativas. Reconociendo la falta de receptividad de muchos interesados frente a los informes largos y demasiado técnicos, el equipo podría elaborar además informes más innovadores y amables con el usuario.
- Amplíe el espectro de interesados que se invita a la presentación y la revisión de los hallazgos e incluya algunos de los grupos de la comunidad y de la sociedad civil con los que a menudo trabajan los evaluadores/as, muchos de los cuales no son consultados en muchas evaluaciones cuantitativas.

Anexo 1

LISTA DE VERIFICACIÓN ESTANDARIZADA PARA EVALUAR LA IDONEIDAD Y LA VALIDEZ DE TODOS LOS DISEÑOS DE EVALUACIÓN	
A. Confirmabilidad	
<i>¿Se basan las conclusiones en evidencia disponible, y está relativamente exenta la investigación de prejuicios del investigador/a?</i>	
1.	<i>¿Se encuentran adecuadamente descritos los métodos y procedimientos empleados en la investigación? ¿Los datos de la investigación se mantienen a disposición para ser re-analizados más adelante?</i>
2.	<i>¿Se han incluido datos para respaldar las conclusiones?</i>
3.	<i>¿El investigador/a ha sido explícito y consciente en lo que se refiere a sus supuestos, valores y prejuicios personales?</i>
4.	<i>¿Se utilizaron los métodos adecuados para controlar los sesgos?</i>
5.	<i>¿Se han considerado hipótesis contradictorias o conclusiones antagónicas?</i>
B. Confiabilidad	
<i>¿El proceso de investigación se ha mantenido consistente, coherente y razonablemente estable a través del tiempo y en relación con todos los investigadores/as y métodos? Si se utilizan diseños emergentes, ¿se encuentran claramente documentados los procesos a través de los cuales fue emergiendo el diseño?</i>	
1.	<i>¿Los hallazgos son confiables, coherentes y replicables en relación con todas las fuentes de datos y a través del tiempo?</i>
2.	<i>¿Se recolectaron datos de todo el abanico de contextos, tiempos y encuestados, etc., apropiados?</i>
3.	<i>¿Contaron todos los trabajadores/as de campo con protocolos de recolección de datos comparables?</i>
4.	<i>¿Se realizaron controles de codificación y calidad, y revelaron éstos una correspondencia adecuada?</i>
5.	<i>¿Coinciden entre sí los relatos de diferentes observadores? Si no lo hacen (lo cual sucede a menudo en las investigaciones CUALI), ¿se reconoce y enfrenta este hecho?</i>
6.	<i>¿Se utilizaron evaluaciones de pares o colegas?</i>
7.	<i>¿Están sujetas las conclusiones a “amenazas a la validez del marco conceptual” [ver la Sección H]? De ser así, ¿se han manejado dichas amenazas adecuadamente?***</i>
8.	<i>¿Se explicitan las normas utilizadas para confirmar las proposiciones, hipótesis, etc.?</i>
C. Credibilidad	
<i>¿Son creíbles los hallazgos para las personas estudiadas y para los lectores/as, y se ha transmitido un retrato adecuado de lo estudiado?</i>	
1.	<i>¿Cuán ricas en contexto y significativas (“sustantivas”) son las descripciones? ¿Hay suficiente información como para ofrecer una descripción creíble/válida de los sujetos o la situación que se está evaluando?***</i>
2.	<i>¿El relato suena verídico, tiene sentido, parece convincente? ¿Refleja el contexto local?</i>
3.	<i>¿La triangulación entre métodos y fuentes de datos complementarios produjo conclusiones en general convergentes? Si se emplean métodos CUALI expansionistas en los que las interpretaciones no necesariamente convergen, ¿se han anotado y analizado las diferencias en las interpretaciones y conclusiones?***</i>
4.	<i>¿Los datos presentados se encuentran adecuadamente vinculados a las categorías de la teoría previa o emergente? ¿Los hallazgos son internamente coherentes, y los conceptos se encuentran sistemáticamente relacionados?</i>
5.	<i>¿Se han identificado áreas de incertidumbre? ¿Se han buscado y encontrado evidencias negativas? ¿Cómo se han utilizado? ¿Se han considerado activamente explicaciones antagónicas?</i>
6.	<i>¿Las conclusiones fueron consideradas acertadas por los investigadores/as responsables de la recolección de datos?</i>
7.	<i>¿Están sujetos los hallazgos a “amenazas a la validez interna” [ver la Sección G]? De ser así, ¿se</i>

manejaron dichas amenazas?***	
8. ¿Están sujetos los hallazgos a “amenazas a la validez estadística”? [ver la Sección F]? De ser así, ¿se manejaron éstas adecuadamente?***	

D. Transferibilidad	
<i>¿Las conclusiones son adaptables a otros contextos y qué tan ampliamente pueden ser generalizadas?</i>	
1. ¿Las características de la muestra de personas, contextos, procesos, etc., están descritas en suficiente detalle como para permitir comparaciones con otras muestras?	
2. ¿En teoría, el diseño de la muestra puede ser generalizado con otras poblaciones?	
3. ¿Define el investigador/a los alcances y límites de una generalización razonable de la investigación?	
4. ¿Los hallazgos incluyen suficientes “descripciones sustantivas” para que los lectores/as evalúen la potencial transferibilidad de la intervención?	
5. ¿Ha constatado una gama de lectores/as que los hallazgos son coherentes con su propia experiencia?	
6. ¿Los hallazgos confirman o coinciden con la teoría existente? ¿Se explicita en el estudio la teoría sobre la transferibilidad del proyecto?	
7. ¿Los procesos y hallazgos son lo suficientemente genéricos como para ser aplicables a otros contextos?	
8. ¿Se han preservado las secuencias narrativas? ¿Se ha formulado una teoría general de casos cruzados utilizando las secuencias?	
9. ¿El informe sugiere contextos en los que los hallazgos podrían ser validados otra vez de manera fructífera?	
10. ¿Los hallazgos han sido replicados en otros estudios con el fin de evaluar su robustez? De no ser así, ¿se podrían organizar con facilidad esfuerzos de réplica?	
11. ¿Están sujetos los hallazgos a “amenazas a la validez externa” [ver la Sección I]? De ser así, ¿de qué forma se han manejado dichas amenazas?	
E. Utilización	
<i>¿Cuán útiles fueron los hallazgos para los clientes, investigadores/as y comunidades estudiadas?</i>	
1. ¿Los hallazgos son intelectual y físicamente accesibles para los potenciales usuarios?	
2. ¿Se hicieron algunas predicciones en la investigación y, de ser así, cuán acertadas fueron éstas?	
3. ¿Los hallazgos proveen orientación para acciones futuras?	
4. ¿Los hallazgos tienen un efecto catalizador conducente a acciones específicas?	
5. ¿Las acciones emprendidas realmente ayudan a resolver los problemas locales?	
6. ¿Los usuarios de los hallazgos han experimentado algún sentido de empoderamiento o de mayor control sobre sus vidas? ¿Han desarrollado nuevas capacidades?	
7. ¿El informe menciona explícitamente consideraciones éticas o basadas en valores? De no ser así, ¿existen algunas consideraciones de este tipo a las que el investigador/a no les esté prestando atención?	

Fuente: Adaptado de Miles y Huberman (1994), Capítulo 10, Sección C (ver también Guba y Lincoln, 1989).

Anexo 2	
LISTA DE VERIFICACIÓN PARA EVALUAR LAS AMENAZAS A LA VALIDEZ DE LOS DISEÑOS DE EVALUACIÓN CUANTITATIVOS	
F. Amenazas a la validez de la conclusión estadística. Razones por las cuales las inferencias sobre la covariación entre dos variables pueden ser incorrectas:	
1. La muestra es demasiado pequeña como para detectar los efectos del programa (bajo poder estadístico): la muestra es demasiado pequeña como para que sea posible detectar diferencias estadísticamente significativas entre los grupos, incluso si éstas existieran.	
2. Supuestos de las pruebas estadísticas vulnerados: Muchas pruebas estadísticas requieren que las observaciones se realicen de manera independiente unas de otras. Si este supuesto es vulnerado, por ejemplo, estudiando a los niños y niñas de un mismo salón de clases, o a pacientes de un mismo consultorio que pueden ser más similares entre sí que la población en general, este hecho puede aumentar el riesgo de error Tipo I (ver el Capítulo 16) y llevar a la conclusión equivocada de que el proyecto tuvo un efecto.	
3. “Andar a la pesca” de resultados estadísticamente significativos: un porcentaje determinado de las pruebas estadísticas arrojará resultados “significativos” por casualidad (uno en 20 si se utiliza el nivel de significación 0.05). Si se “anda a la pesca” en una gran cantidad de cuadros estadísticos, siempre se detectará algunos de estos resultados espurios.	
4. Poca confiabilidad de las mediciones de cambios en los indicadores de efectos directos. Las mediciones poco confiables de, por ejemplo, las tasas de cambio en el ingreso, alfabetismo o mortalidad infantil, siempre reducen la probabilidad de detectar un efecto directo significativo.	
5. Restricción del rango. Si se establecen comparaciones entre grupos similares solamente se reduce el poder de la prueba, y también la probabilidad de detectar un efecto significativo.	
6. Poca confiabilidad de la ejecución del tratamiento. Si no se aplica el tratamiento de manera idéntica a todos los sujetos, se reduce la probabilidad de detectar un efecto significativo.	
7. Eventos externos influyen en los efectos directos (varianza extraña en un contexto experimental). Eventos o presiones de naturaleza externa (cortes de electricidad, violencia comunitaria, campañas electorales) pueden distraer la atención de los sujetos y afectar conductas y los efectos directos del programa.	
8. Diversidad de la población (heterogeneidad de las unidades). Si los sujetos tienen características considerablemente distintas, este hecho puede incrementar la varianza de los resultados y dificultar la detección de efectos significativos.	
9. Estimación inexacta del tamaño del efecto. Unas cuantas variables con valores muy por encima o por debajo del rango esperado (valores extremos) pueden reducir significativamente el tamaño del efecto (ver el Capítulo 14) y reducir la probabilidad de que se detecten diferencias significativas.	
10. Extrapolación de una base de datos truncada o incompleta**. Si la muestra cubre solo una parte de la población (por ejemplo, solo las familias más pobres, o solo las personas que trabajan en el sector formal), este hecho puede afectar las conclusiones del análisis y generar generalizaciones sesgadas acerca de la población total.	
11. Las muestras de grupos del proyecto y de comparación no cubren las mismas poblaciones**. A menudo se da el caso de que la muestra del grupo de comparación no se extrae de una población con características idénticas a las de la muestra del proyecto. En estos casos las diferencias en los efectos directos pueden obedecer a las diferencias en las características de las dos muestras, y no a los efectos directos del proyecto.	
12. No se recolecta información de las personas adecuadas, o no se entrevista a algunas categorías de informantes**. En ocasiones se recoge información únicamente de y sobre ciertos sectores de la población objetivo (hombres y no mujeres, maestros/as y no alumnos/as, etc.), en cuyo caso los estimados referidos a la población total pueden ser sesgados.	

<p>G. Amenazas a la validez interna. Razones por las cuales inferir que la relación entre dos variables es causal puede ser incorrecto.</p>	
<p>1. No está claro si la intervención del proyecto ocurrió realmente antes del efecto presumido (precedencia temporal ambigua). Una causa obligatoriamente tiene que ser anterior a su efecto. Sin embargo, muchas veces es difícil determinar el orden de los eventos en un proyecto. Muchos proyectos (por ejemplo, programas de desarrollo urbano) no cuentan con una fecha de inicio precisa sino que empiezan a ejecutarse a lo largo de varios meses o incluso años.</p>	
<p>2. Selección. Los participantes del proyecto a menudo son diferentes a los grupos de comparación, ya sea porque se seleccionan a sí mismos o porque el proyecto selecciona a personas con ciertas características (los agricultores más pobres o las comunidades mejor organizadas).</p>	
<p>3. <i>Historia.</i> La participación en un proyecto puede generar otras experiencias no relacionadas con el tratamiento del proyecto que distingue a los grupos del proyecto y de control. Por ejemplo, los empresarios conocidos por haber recibido préstamos pueden ser más propensos a que les roben o a recibir presiones de los políticos para que efectúen donaciones, o las niñas matriculadas en secundaria pueden ser más propensas a quedar embarazadas.</p>	
<p>4. Maduración. El proceso de maduración produce muchos cambios naturales en el comportamiento, el conocimiento y la exposición a nuevas experiencias. A menudo es difícil separar los cambios que tienen su origen en la maduración de aquellos que se deben efectivamente al proyecto.</p>	
<p>5. Regresión a la media. Si se selecciona a los sujetos en razón de sus puntajes extremos (peso, desarrollo físico), existe una tendencia natural a desplazarse más cerca de la media con el tiempo – disminuyendo o distorsionando con ello los efectos del programa.</p>	
<p>6. Atrición. Incluso cuando los participantes del proyecto originalmente tenían características similares a las de la población total, la deserción selectiva a través del tiempo puede modificar las características de la población del proyecto (por ejemplo, es posible que deserten los más pobres o los menos educados).</p>	
<p>7. Pruebas. Ser entrevistado o sometido a una prueba puede afectar el comportamiento o las respuestas. Por ejemplo, una persona a la que se le hacen preguntas sobre sus gastos puede declarar que gasta menos en artículos que no son socialmente aceptados (cigarrillos o alcohol) y más en artículos aceptables.</p>	
<p>8. Los investigadores/as pueden cambiar la forma en que describen y/o interpretan los datos a medida que adquieren más experiencia (instrumentación). A medida que los entrevistadores/as u observadores/as adquieren mayor experiencia, pueden variar la forma en que interpretan las escalas de calificación, las listas de verificación de observaciones, etc.</p>	
<p>9. Distorsión por parte del encuestado al utilizar el método de retrospectiva**. Los encuestados pueden distorsionar, deliberadamente o no, sus recuerdos de eventos pasados. Políticos de oposición pueden exagerar los problemas de la comunidad, mientras que los ancianos de la comunidad pueden recordar el pasado con un velo de romanticismo.</p>	
<p>10. Uso de diseños menos rigurosos por razones presupuestarias y de tiempo**. Las EMR son especialmente vulnerables a una gran variedad de problemas de validez interna, debido a la necesidad de eliminar costos y ahorrar tiempo.</p>	

H. Amenazas a la validez del marco conceptual. Razones por las cuales las inferencias acerca de los marcos conceptuales que caracterizan las operaciones del estudio pueden ser incorrectas	
1. Explicación inadecuada de los marcos conceptuales. Los marcos conceptuales (los efectos / efectos directos) que están siendo estudiados se encuentran definidos en términos ya sea muy generales o confusos / ambiguos, imposibilitando con ello una medición exacta (ejemplos de marcos conceptuales ambiguos: desempleados, agresivos, ambiente laboral hostil, discriminación sexual).	
2. Los indicadores no miden adecuadamente los marcos conceptuales (confusión en torno al marco conceptual). La definición operativa puede no reflejar adecuadamente el marco conceptual deseado. Por ejemplo, definir a los desempleados como las personas que se han registrado en un centro de empleo pasará por alto a muchas personas que se encuentran desempleadas pero no acuden a dichos centros. Del mismo modo, definir la violencia doméstica como los casos denunciados a la policía implica una subrepresentación notable del número real de incidentes.	
3. Uso de un solo indicador para medir un marco conceptual complejo (sesgo de mono-operación). Usar un solo indicador para definir y medir un marco conceptual complejo (por ejemplo, la pobreza, el bienestar, la violencia doméstica) suele dar lugar a sesgos.	
4. Uso de un solo método para medir un marco conceptual (sesgo de mono-método). Si se emplea un solo método para medir un marco conceptual, la medición será estrecha y a menudo sesgada (por ejemplo, observar a las comunidades en asambleas formales producirá resultados distintos que observarla en eventos sociales o faenas comunitarias como parte de un proyecto).	
5. Estudio de un solo nivel de tratamiento (confundir los marcos conceptuales con niveles de marcos conceptuales). Muchas veces se aplica un tratamiento a un solo nivel de intensidad, y éste suele ser bastante bajo (solo préstamos a pequeñas empresas, solo un programa de capacitación corto para dirigentes comunitarios, entrega de solo unas cuantas computadoras a una escuela) y los resultados se utilizan para extraer conclusiones generales acerca de la eficacia (o la falta de ella) del marco conceptual. Esto lleva a conclusiones erróneas, dado que un nivel de tratamiento más alto podría haber producido un efecto más significativo.	
6. Los participantes del programa y el grupo de comparación responden a algunas preguntas de manera distinta (estructura factorial sensible al tratamiento). Los participantes del programa pueden ofrecer respuestas que incluyen más matices a las mismas preguntas. Por ejemplo, pueden hacer una distinción entre diferentes tipos e intensidades de violencia doméstica o prejuicios raciales, mientras que el grupo de comparación puede ofrecer respuestas más generales y menos discriminadas.	
7. Los participantes se evalúan a sí mismos y su situación de manera diferente que el grupo de comparación (cambios reactivos en los autoinformes). Las personas seleccionadas para un programa pueden informar sobre sí mismas de manera diferente que aquellas que no han sido seleccionadas, incluso antes del inicio del programa. Posiblemente deseen crear la impresión de que necesitan el programa con mayor urgencia (que son más pobres o están más enfermas) o que tienen más méritos si se tratara de un criterio de selección.	
8. Reactividad a la situación experimental. Los participantes del proyecto intentan interpretar la situación del proyecto, lo cual puede afectar su comportamiento. Si creen que el proyecto es dirigido por una organización religiosa o un grupo político en particular, pueden ofrecer las respuestas que consideran que se esperan de ellos.	
9. Expectativas del experimentador/a. Los experimentadores/as también albergan expectativas (sobre la reacción de hombres y mujeres de diferentes grupos socioeconómicos al programa) que pueden afectar su reacción frente a dichos grupos.	
10. Efectos de la novedad y la interrupción. Los programas novedosos pueden generar excitación y producir un efecto importante. Si un programa similar es introducido posteriormente, el efecto puede ser menor debido a que ha dejado de ser novedoso.	
11. Efectos compensatorios y rivalidad. Los programas generan una dinámica que puede afectar los efectos directos a una serie de niveles. Puede haber presiones (políticas u otras) para otorgar beneficios a sectores que no participan en el proyecto; puede surgir una rivalidad con	

<p>los grupos de comparación, que posiblemente intenten demostrar que pueden lograr los mismos resultados por su cuenta; o los sectores que no reciben ningún tratamiento o reciben uno menos atractivo pueden sentirse desmoralizados.</p>	
<p>12. Uso de indicadores y marcos conceptuales provenientes de otros países sin realizar pruebas en el contexto local**. Muchas evaluaciones utilizan importantes teorías y marcos conceptuales provenientes de otros países, que posiblemente no recojan adecuadamente la situación del lugar donde se desarrolla el proyecto. Por ejemplo, muchas evaluaciones de los impactos de programas de microcrédito en el empoderamiento de las mujeres en países como Bangladesh e India han utilizado definiciones de empoderamiento que posiblemente no sean apropiadas para las mujeres de Bangladesh.</p>	

<p>I. Amenazas a la validez externa. Razones por las cuales pueden ser incorrectas las inferencias en el sentido que los resultados del estudio serian validos con diferentes poblaciones, contextos, tratamientos y efectos directos.</p>	
<p>1. La muestra no cubre a toda la población de interés. Los sujetos pueden ser de un solo sexo o de grupos étnicos o económicos específicos o pueden tener determinadas características de personalidad (por ejemplo, pueden estar deprimidos o ser seguros de sí mismos, etc.). En consecuencia, puede ser difícil hacer generalizaciones aplicables a toda la población a partir de los resultados del estudio.</p>	
<p>2. Los efectos directos del programa se ven afectados por contextos diferentes (interacción de la relación causal con variaciones en el tratamiento). Los tratamientos pueden ser aplicados en contextos distintos, lo cual posiblemente afectará los efectos directos. Si presiones de reducir el tamaño de las clases obliga a las escuelas a construir aulas adicionales, provisionales e inadecuadas, los efectos directos pueden ser muy diferentes que si se cuenta con clases más pequeñas en el contexto de salones de clase apropiados.</p>	
<p>3. Distintas mediciones de los efectos directos arrojan distintas evaluaciones de la eficacia del proyecto (interacción de la relación causal con los efectos directos). Mediciones distintas de los efectos directos pueden producir conclusiones distintas sobre la eficacia del proyecto. Los programas de microcrédito para mujeres pueden incrementar el ingreso del hogar y el gasto en la educación de los hijos pero posiblemente no aumenten el empoderamiento político de la mujer.</p>	
<p>4. Los efectos directos del programa pueden variar en contextos distintos (interacciones de las relaciones causales con los contextos). El éxito de un programa puede ser diferente en contextos rurales y urbanos o en comunidades de distinta naturaleza. En consecuencia, es posible que sea inapropiado generalizar los hallazgos de un contexto haciéndolos extensivos a otros.</p>	
<p>5. Los programas funcionan de manera diferente en contextos distintos (mediación dependiente del contexto). Los programas pueden funcionar de manera distinta y tener efectos directos intermedios y finales distintos en contextos diferentes. La implementación de escuelas gestionadas por la comunidad puede funcionar de manera muy diferente y tener efectos directos distintos si es manejada por organizaciones religiosas, instituciones estatales u organizaciones no gubernamentales.</p>	
<p>6. La actitud de quienes diseñan las políticas y los políticos frente al programa**. Programas idénticos funcionarán de manera diferente y tendrán efectos directos distintos si cuentan con el apoyo activo de quienes diseñan las políticas y los políticos, en comparación con situaciones en las que enfrentan la oposición o la indiferencia de dichos actores. Cuando se producen cambios en el partido gobernante o en la dirigencia de una organización, es común que desaparezca o bien se fortalezca el apoyo a los programas.</p>	
<p>7. Ciclos estacionales y de otra índole**. Muchos proyectos funcionan de manera diferente en diferentes estaciones, en distintas etapas del ciclo economico o de acuerdo a los términos de comercio que rigen para exportaciones e importaciones clave. Los intentos de generalizar los hallazgos de un programa piloto deben tomar en cuenta estos criterios.</p>	

Fuente: Adaptado de Shadish, Cook y Campbell (2002), Tablas 2.2, 2.4, 3.1 y 3.2.

** Estas amenazas han sido añadidas por los autores/as de este informe por revestir una pertinencia especial para los escenarios de Evaluación en un Mundo Real.

Referencias

[Ver una lista referencial más exhaustiva en Bamberger, Rugh y Mabry, 2006. *RealWorld Evaluation*, Sage Publications]

- Ahmed, Nizam. *Study on gender dimensions of the Second Bangladesh Rural Roads and Markets Improvement and Maintenance Project*. Disponible en www.worldbank.org/genddr/transport. Hacer clic en “Grants and Pilots”.
- Aron, A., y Aron, E., 2002. *Statistics for the Behavioral and Social Sciences*. Prentice Hall.
- Bamberger, M. (editor), 2000. *Integrating Quantitative and Qualitative Research in Development Projects*. Directions in Development Series, Banco Mundial.
- Banco Mundial, Departamento de Evaluación de Operaciones, 2004. *Influential Evaluations*. Banco Mundial, Departamento de Evaluación de Operaciones, 2005. *Influential Evaluations: Detailed Case Studies*. Disponible sin costo alguno en: www.worldbank.org/ieg/ecd.
- Banco Mundial, Grupo de Evaluación Independiente, 2006. Conducting quality impact evaluations under budget, time and data constraints. Disponible sin costo alguno en: www.worldbank.org/ieg/ecd.
- Brewer, J., y Hunter, A., 2006 (editores). *Foundations of Multimethod Research. Synthesizing Styles*, Sage Publications.
- Brown, J., 2000. “Evaluating the impact of water supply projects in Indonesia”. En Bamberger, M. (editor), *Integrating Quantitative and Qualitative Research in Development Projects*. Directions in Development Series, Banco Mundial.
- Creswell, J.W., Plano Clark, V.L., Guttman, M.L., y Hanson, W., 2003. Advanced Mixed Method Research Designs. En Tashakkori, A., y Teddue, C. (editores), *Handbook of Mixed Methods in Social and Behavioral Science* (pp. 209-240). Thousand Oaks, CA, Estados Unidos: Sage .
- Meyers, L., Gamst, G., y Guarino, A., 2006. *Applied Multivariate Research: Design and Interpretation*. Sage Publications.
- Glewwe, P., Kremer, M., Sylvie Moulin y Zitzewitz, E., 2004. Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya. *Journal of Development Economics* (74: 251-268).
<http://www.povertyactionlab.com/projects/project.php?pid=26>
- Kertzner, D., y Fricke, T., 1997. “Toward an anthropological demography”. En “*Anthropological demography - Towards a new synthesis*” (editor), Kertzner and Fricke, University of Chicago Press.
- Khandker, S., 1998. *Fighting Poverty with Microcredit: Experience in Bangladesh*. Oxford University Press.
- Kozel, V., y Parker, B., 2000. “Integrated approaches to poverty assessment in India”. En Bamberger, M. (editor), 2000. *Integrating Quantitative and Qualitative Research in Development Projects. Directions in Development* (pp. 59-68). Washington D.C, Estados Unidos: Banco Mundial.
- Kumar, S., 2002. *Methods for Community Participation. A Complete Guide for Practitioners*. London, ITDG Publishing.

- Mabry, L., 1998. "Case study methods" (pp. 155-70). En *Evaluation Research for Educational Productivity*. Editado por H.J. Walberg y A.J. Reynolds, Greenwich, CT, JAI Press.
- Newman, C., 2001. Gender, time use, and change: impacts of agricultural export employment in Ecuador. *Policy Research Report on Gender and Development Working Paper Series No. 18*. Poverty Reduction and Economic Management Network/Development Research Group. Banco Mundial, febrero de 2001. Disponible en la página web www.worldbank.org/gender/prr
- Patton, M.Q., 1997 (tercera edición). *Utilization-focused evaluation*. Thousand Oaks, California: Sage Publications.
- Patton, M.Q., 2002. *Qualitative research and evaluation methods*. Thousand Oaks, California: Sage Publications.
- Pradhan, Menno, y Laura Rawlings, 2000. "The impact of targeting of social infrastructure investments: lessons from the Nicaraguan Social Fund." *World Bank Economic Review*, 16(2): 275-295.
- Presser, P., Couper, M., Lessler, J., Martin, M., Martin, J., Rothgeb, J., y Singer, E. Methods for Testing and Evaluating Survey Questionnaires. New York: Wiley.
- Presser, P., Couper, M., Lessler, J., Martin, M., Martin, J., Rothgeb, J., y Singer, E. 2004. Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly*, Vol. 68, No. 1, pp. 109-130.
- Ravallion, M., 2006. Evaluating anti-poverty programs. *Handbook for Agricultural Economics* (editado por Robert Evenson y Paul Schulz). Vol. 4, North-Holland.
- Rietberger-McCracken, J., y Narayan, D., 1997. Participatory Rural Appraisal. Módulo III de *Participatory tools and techniques: a resource kit for participation and social assessment*. Departamento de Medio Ambiente, Washington D.C: Banco Mundial.
- Roche, C., 1999. *Impact Assessment for Development Agencies*. Learning to Value Change, OXFAM.
- Rugh, J., 1986. *Self-Evaluation: Ideas for Participatory Evaluation of Rural Community Development Projects*. Oklahoma City, OK: World Neighbors.
- Schwarz, N., y Oyserman, D., 2001. Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation*, Vol. 22, No. 2, pp.127-160.
- Sirken, R., 1999. *Statistics for the Social Sciences*. Thousand Oaks. Sage Publications.
- Tashakkori, A., y Teddue, C. (editores), 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA, Estados Unidos, Sage Publications.
- Vaughan, R., y Buss, T., 1994. *Communicating Social Science Research to Policymakers*. Applied Social Science Research Methods Series No. 48, Sage Publications.
- Valadez, J., y Bamberger, M., 1994. *Monitoring and evaluating social programs in developing countries: a handbook for policymakers, managers and researchers*. Washington D.C., Estados Unidos: Banco Mundial.