



Evaluación Corporativa

Las Evaluaciones de Impacto del BID: Producción, Uso e Influencia

Las Evaluaciones de Impacto del BID: Producción, Uso e Influencia

Oficina de Evaluación y Supervisión (OVE)
Octubre 2017





Este trabajo se distribuye bajo la licencia de Creative Commons https://creativecommons.org/licenses/by-ncnd/3.0/us/deed.es_ES (CC BY-NC-ND 3.0 US). Usted es libre de compartir, copiar y redistribuir el material en cualquier medio o formato bajo las siguientes condiciones:



Reconocimiento – Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.



No comercial - No puede utilizar el material para una finalidad comercial.



Sin obras derivadas - Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

No hay restricciones adicionales – No puede aplicar términos legales o medidas tecnológicas que legalmente restrinjan realizar aquello que la licencia permite.

El enlace URL incluye términos y condicionales adicionales de esta licencia.

© **Banco Interamericano de Desarrollo, 2017**

Oficina de Evaluación y Supervisión
1350 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org/evaluacion

RE-512-1
Octubre 2017

SIGLAS Y ABREVIATURAS

RECONOCIMIENTOS

RESUMEN EJECUTIVO

1. ANTECEDENTES	1
2. LA PRODUCCIÓN DE EVALUACIONES DE IMPACTO EN EL BID	9
A. Análisis de la producción	9
1. ¿Cuántas evaluaciones de impacto se han propuesto y en qué situación se encuentran?	9
2. ¿Cómo han funcionado los procesos de originación, diseño e implementación?	16
3. ¿Cuánto cuestan las evaluaciones de impacto?	21
4. ¿Cuáles son las buenas prácticas en materia de financiamiento de las evaluaciones de impacto y los productos de conocimiento en general?	25
B. ¿Cuál es el nivel de calidad de las evaluaciones completadas y del diseño de las que están en marcha?	27
3. USO E INFLUENCIA DE LAS EVALUACIONES DEL BID	33
A. Uso e influencia de las evaluaciones de impacto como producto de conocimiento	33
1. ¿En qué medida se usan las evaluaciones de impacto en el BID?	33
2. ¿Cómo emplean los clientes las evaluaciones de impacto del BID?	36
3. ¿Cuál es el alcance más amplio de las evaluaciones de impacto del BID?	39
B. ¿Pueden las evaluaciones de impacto facilitar la ejecución de proyectos?	41
4. CONCLUSIONES Y RECOMENDACIONES	45

NOTAS

REFERENCIAS Y BIBLIOGRAFÍA

Disponibles en www.iadb.org/evaluation

ANEXO I	Creación de la base de datos de evaluaciones de Impacto del BID
ANEXO II	Estadísticas básicas sobre de las evaluaciones de impacto en el BID
ANEXO III	Las evaluaciones de impacto en otras organizaciones (DIME, J-PAL y Millennium Challenge Corporation)
ANEXO IV	Prácticas óptimas en materia de financiamiento del conocimiento
ANEXO V	Evaluación de CLEAR de la calidad de las evaluaciones de impacto del BID
ANEXO VI	Uso de pruebas empíricas en la formulación de políticas en América Latina y el Caribe: el caso de las evaluaciones de impacto del BID
ANEXO VII	¿Facilitan las evaluaciones de impacto la ejecución de los proyectos del BID?
ANEXO VIII	Análisis del uso de las evaluaciones de impacto en los documentos de marco sectorial

CCB	Departamento de Países del Grupo Caribe
CID	Departamento de Países de Centroamérica, México, Panamá y la República Dominicana
CLEAR-LAC	Centro para el Aprendizaje en Evaluación y Resultados para América Latina y el Caribe
CSD	Sector de Cambio Climático y Desarrollo Sostenible
DEM	Matriz de Efectividad en el Desarrollo
DIME	Evaluación del Impacto sobre el Desarrollo
EDU	División de Educación
GDI	División de Género y Diversidad
IFD	Sector de Instituciones para el Desarrollo
INE	Sector de Infraestructura y Energía
INT	Sector de Integración y Comercio
J-PAL	Abdul Latif Jameel Poverty Action Lab
KNL	Sector de Conocimiento y Aprendizaje
RES	Departamento de Investigación y Economista Jefe
RND	División de Medio Ambiente, Desarrollo Rural y Administración de Riesgos por Desastres
SCL	Sector Social
SFD	Oficina de Planificación Estratégica y Efectividad en el Desarrollo
SPH	División de Protección Social y Salud
VPC	Vicepresidencia de Países
VPS	Vicepresidencia de Sectores y Conocimientos

Este documento fue elaborado bajo la orientación de Cheryl Gray (Directora de OVE) por un equipo integrado por Anna Crespo y Oliver Azuara (jefes de equipo), Cesar Bouillon, Veronica Gonzalez Diez, Alejandro Palomino, Florencia Alejandro, Johanan Rivera y Maya Jansson.

OVE agradece al personal del BID por sus constructivos comentarios durante el proceso de evaluación y también a las personas de otras instituciones y organismos ejecutores que participaron en encuestas y entrevistas en el contexto del presente estudio.



La adopción de la Matriz de Efectividad en el Desarrollo ha incentivado considerablemente la producción y el uso de las evaluaciones de impacto.

© BID

Resumen Ejecutivo

Durante el último decenio, el Banco Interamericano de Desarrollo (BID o el Banco) ha invertido considerablemente en la promoción de la evaluación de impacto como herramienta para medir la eficacia de la institución. Al favorecer el uso de evaluaciones de impacto, la intención del Banco ha sido fomentar intervenciones más eficaces. La adopción de la Matriz de Efectividad en el Desarrollo (DEM), el primer esfuerzo emprendido a nivel de todo el Banco para promover proyectos más evaluables, ha incentivado considerablemente la producción y el uso de dichas evaluaciones. La DEM incluye una serie de requisitos de seguimiento y evaluación, cuyo cumplimiento sirve para calificar las operaciones. Esta evolución ha ocurrido en el contexto de un cambio cultural general hacia la evaluación y análisis más rigurosos, que son componentes fundamentales del Noveno Aumento General de Recursos y la nueva Estrategia Institucional. Es en ese contexto que OVE evaluó la producción y el uso de las evaluaciones de impacto en las operaciones del BID del sector público, procurando hacer un balance y examinar los frutos que ese esfuerzo ha traído para el Banco, sus costos y el rumbo que el BID debe tomar en los próximos años.

El Banco no mantiene una base de datos centralizada, un registro oficial o una clasificación presupuestaria de las evaluaciones de impacto; no existe un mecanismo institucional que permita determinar fácilmente su producción. Asimismo, como se preparan en muchos contextos diferentes, es difícil identificar las evaluaciones de impacto en función de actividades o registros¹. OVE creó una base de datos para analizar

cada una de las evaluaciones de impacto propuestas por el Banco en sus documentos de préstamo y perfiles de cooperación técnica, y recabó información adicional por medio de análisis documentales y amplias entrevistas con el personal del BID, otros bancos multilaterales de desarrollo, clientes y expertos en evaluaciones de impacto.

Entre 2006 y 2016 el Banco propuso 531 evaluaciones de impacto en documentos de préstamo y proyectos de cooperación técnica. De esa cifra, 94 se han completado, 286 están en marcha (en diferentes etapas) y 151 se han cancelado por motivos como la cancelación del proyecto, cambios políticos y problemas de diseño e implementación, lo que incluye la falta de una metodología bien definida para evaluación, en la fase de aprobación. El personal del BID diseña, en promedio, el 60% de esas evaluaciones, porcentaje que alcanzó el 80% en 2016. De las evaluaciones que se propusieron en los documentos de préstamo, alrededor del 34% se emprendió principalmente debido al puntaje en la DEM y estas eran las que tenían mayor probabilidad de cancelarse. De lo anterior puede concluirse que el compromiso de los gobiernos y el interés genuino en aprender y utilizar los resultados hacen que las evaluaciones de impacto sean más exitosas.

El análisis de originación muestra una continua falta de priorización en la selección de las evaluaciones de impacto. De una revisión minuciosa de los 20 documentos de marco sectorial vigentes se desprende que solo alrededor del 30% incluye una revisión exhaustiva de la literatura especializada en que se identifican brechas de conocimiento y se definen ámbitos específicos de trabajo analítico que podrían orientar la priorización de las evaluaciones de impacto.

Aunque los costos de las evaluaciones del BID varían bastante, son ampliamente comparables a los de otros organismos y se ajustan a los parámetros de referencia señalados en la literatura especializada. El principal factor que impulsa el costo de las evaluaciones de impacto es la necesidad de recopilar datos: las evaluaciones que exigen la recopilación de datos cuestan en promedio US\$468.000, en tanto que las que emplean datos administrativos cuestan en promedio menos de US\$80.000. Si bien estos montos excluyen el tiempo de los especialistas del Banco, los datos que se recopilaron para realizar la presente evaluación indican que la mayor parte de los costos corresponden a la recopilación de datos y los consultores. En promedio, el costo de una evaluación de impacto para un proyecto es menos del 0,5% del monto del préstamo.

El Banco presupuestó alrededor de US\$200 millones para evaluaciones de impacto durante el período de 2006 a 2016, pero aproximadamente el 25% de esos recursos se destinó a evaluaciones que fueron canceladas. De ese total, US\$152 millones se sufragaron mediante operaciones, tanto de préstamo como de financiamiento no reembolsable, y los US\$54 millones restantes mediante operaciones de cooperación técnica (estas últimas no se ciñeron a ninguna estrategia general). Si solo se toman en cuenta las evaluaciones de impacto en los países en los que no se realizaron operaciones de financiamiento no reembolsable para inversión con dichas evaluaciones, el BID gastó aproximadamente US\$0,61 de sus recursos del Capital Ordinario y de fondos fiduciarios por cada US\$1 que los países destinaron al financiamiento de esas evaluaciones.

La Oficina de Evaluación y Supervisión (OVE) identificó numerosas buenas prácticas en los bancos multilaterales de desarrollo con respecto al financiamiento de la producción de conocimiento. Un proceso de priorización transparente y estratégico permite a los bancos multilaterales de desarrollo centrar la atención en las brechas de conocimiento en sectores específicos, facilita el apalancamiento de los recursos y reduce los costos de poner en práctica el conocimiento y difundirlo fomentando relaciones de largo plazo. Los modelos exitosos de financiamiento de evaluaciones de impacto muchas veces incluyen la participación activa de los usuarios finales del conocimiento. Dado que el conocimiento es un bien público, la cooperación entre las redes facilita la producción de evaluaciones de impacto, así como su pertinencia y visibilidad. Por último, la mayor parte de los organismos aplican protocolos de investigación claramente definidos, lo que ayuda a asegurar que los hallazgos se publiquen y atrae la participación de donantes.

La calidad de las evaluaciones de impacto del Banco parece haber mejorado con el tiempo, pero sigue habiendo margen para mejoras. OVE contrató al Centro para el Aprendizaje en Evaluación y Resultados para América Latina y el Caribe (CLEAR-LAC) para que examinara la calidad de las evaluaciones de impacto completadas y de una muestra de propuestas de evaluación. Los criterios que empleó CLEAR se ajustaron a las normas internacionales y abarcaron la pertinencia de la pregunta de evaluación, la idoneidad de los datos empleados, el rigor de la metodología y la solidez del análisis. Los especialistas de CLEAR concluyeron que la calidad de alrededor del 55% de las evaluaciones de impacto completadas era satisfactoria o parcialmente satisfactoria. Si bien esa calidad disminuyó tras la adopción de la DEM debido a que otros sectores que normalmente no realizaban evaluaciones de impacto también empezaron a prepararlas, luego el nivel siguió aumentando. Asimismo, la calidad de las propuestas también ha seguido una clara tendencia de mejora.

Independientemente de su calidad, las evaluaciones de impacto solo pueden ejercer influencia si se emplean, y en ese sentido OVE constató que la accesibilidad sigue siendo un factor. De las evaluaciones completadas, 55 (58%) no se divulgaron ni en publicaciones especializadas ni en las series de documentos de trabajo del BID. Muchos ni siquiera pudieron encontrarse en IDBDOCS, y OVE tuvo que pedírselas directamente al jefe de equipo. Ello demuestra que los resultados de las evaluaciones de impacto no son de fácil acceso para el público. No obstante, el análisis de citas muestra que el uso de las evaluaciones de impacto del Banco ha ido en aumento, lo que ha ayudado al BID a ganar un poco de reputación como banco de conocimiento.

El análisis de los documentos del Banco indica que las pruebas empíricas derivadas de las evaluaciones de impacto no se mencionan con frecuencia, aunque algunos clientes parecen estar usando las evaluaciones más directamente. Los clientes tuvieron diferentes puntos de vista sobre el enfoque del Banco con respecto a las evaluaciones de impacto, pero solo 2 de las 25 entrevistas realizadas fueron muy negativas. Si bien los demás entrevistados señalaron deficiencias en el proceso, la mayoría estuvo satisfecha con los resultados, ya sea porque la evaluación de impacto ayudó a asegurar la sostenibilidad o continuidad de los proyectos o porque creó capacidad para seguir realizando ese tipo de evaluaciones.

Un uso importante de las evaluaciones de impacto en el Banco es evitar programas ineficaces. Aunque hubo casos en que las evaluaciones apuntaron a resultados nulos y el Banco optó por seguir adelante con el mismo tipo de programa, en general las evaluaciones de impacto parecen haberse utilizado para evitar programas ineficaces. Es probable que ese sea uno de los efectos de mayor alcance de las evaluaciones de impacto del BID, ya que el Banco ayuda a los gobiernos a que eviten destinar recursos a programas que no alcanzarán los resultados deseados.

OVE también encontró algunas pruebas empíricas que apuntan a que las evaluaciones de impacto bien fundamentadas pueden ayudar a respaldar la implementación de proyectos. Si bien las curvas de desembolso de proyectos con y sin evaluaciones de impacto son casi iguales durante los dos primeros años, posteriormente los que sí incluyen evaluaciones de impacto tienen desembolsos más rápidos y se completan tres meses antes. El análisis econométrico indica que estos resultados son robustos cuando se neutralizan varias variables. Aunque no está claro cuáles son exactamente los canales a través de los cuales las evaluaciones de impacto influyen sobre la implementación de proyectos, hay algunas pruebas empíricas que indican que es posible que el respaldo y la capacitación que proporcionan los investigadores y el personal en el terreno que da seguimiento al proyecto influyan en ello. Así pues, las implicaciones financieras de ese mejor desempeño posiblemente podrían contrarrestar hasta la mitad del costo de la evaluación de impacto.

Sobre la base de los hallazgos de la presente evaluación, OVE formula las siguientes recomendaciones para mejorar la producción y el uso de las evaluaciones de impacto en el BID:

1. Adoptar un enfoque más estratégico con respecto a la selección de las evaluaciones de impacto realizando o apoyando solo aquellas que cuenten con el compromiso del cliente (o una estrategia clara para lograr ese compromiso), una brecha de conocimiento bien definida y un cronograma factible. La presente evaluación demuestra la importancia de contar con el apoyo del cliente al realizar una evaluación de impacto. El compromiso del cliente también ayuda a asegurar que los resultados de las evaluaciones de impacto se empleen una vez concluidas.
2. Asegurar que los documentos de marco sectorial señalen las brechas de conocimiento para así ayudar a orientar la labor del Banco en materia de evaluaciones de impacto. A la fecha, los documentos de marco sectorial han hecho esto en distintas medidas, y OVE recomienda que de ahora en adelante lo hagan de forma más sistemática e integral.
3. Revisar la DEM a fin de reducir los incentivos que hacen que se proponga un número excesivo de evaluaciones de impacto. Todos los proyectos deben contar con un sistema sólido de seguimiento y evaluación, pero no todos deben incluir una evaluación de impacto. La DEM ha sido una herramienta útil para incrementar la evaluabilidad de los proyectos y promover una mayor comprensión

de la importancia de realizar evaluaciones más rigurosas. También ha incrementado los incentivos para proponer evaluaciones de impacto, lo que ha hecho que se generen propuestas que no estaban bien pensadas o que no se justificaban.

4. Crear un mecanismo de financiamiento transparente en el que los intereses de los clientes se correspondan con los del Banco. Cuando se considere pertinente y factible realizar una evaluación de impacto para una operación de préstamo, la evaluación debe incluirse en el sistema de seguimiento y evaluación que financie el préstamo en la medida en que resulte informativo para el país y para el informe de terminación de proyecto (ITP). Si hay efectos a más largo plazo que solo pueden medirse una vez se cierre el proyecto, debe financiarse una labor de evaluación ex post con recursos administrados por el Banco. De esa forma, los clientes comprometidos a realizar evaluaciones de impacto y a establecer mecanismos de seguimiento y evaluación adecuados podrán beneficiarse de recursos del BID para completar la evaluación después de que cierre el proyecto. Al mismo tiempo, el Banco comprometería los recursos propios que administra únicamente para las evaluaciones a más largo plazo en que haya un compromiso del cliente, una buena gestión inicial (con parámetros de referencia y seguimiento adecuados) y la expectativa de poder recabar información valiosa sobre los resultados del programa a más largo plazo.
5. Reforzar los sistemas de control de calidad. El Banco ya cuenta con una serie de mecanismos para el control de calidad ex ante. Sin embargo, conforme se muestra en la presente evaluación, no todas las evaluaciones de impacto se completan con un alto nivel de calidad. Para evitar el riesgo reputacional de publicar y difundir trabajos de baja calidad, el Banco necesita crear un sistema de control de calidad que vaya más allá de las evaluaciones de impacto que se publican como documentos de trabajo o notas técnicas.
6. Fortalecer y centralizar los sistemas del Banco para dar seguimiento a todas las evaluaciones de impacto que están en marcha —desde la fase de diseño hasta la final— y los costos y poner al alcance del público las evaluaciones de impacto de alta calidad y su respectiva documentación, los datos empleados (de ser posible) y las lecciones aprendidas. Como se señala en la presente evaluación, SPD y KNL han avanzado en este sentido, y el sistema ha de finalizarse.
7. Estructurar un mecanismo más formal para promover la formación de asociaciones que aumenten la calidad, la pertinencia y la visibilidad de la labor del BID, al tiempo que se reduzcan los costos. La presente evaluación detectó pocos casos en que se crearon asociaciones para elaborar las evaluaciones de impacto en sí. Sobre la base de las buenas prácticas observadas en el BID y otros organismos, podría ser de utilidad que el Banco estableciera más asociaciones con círculos académicos e intensificara su colaboración con otros organismos en la producción y difusión de las evaluaciones de impacto y en el intercambio de información. SPD ha estado promoviendo esas actividades, y ese esfuerzo ha de fortalecerse.

¹ Las evaluaciones de impacto pueden producirse de cuatro formas diferentes: (i) como productos de investigación pura, (ii) como insumos durante la preparación de proyectos, (iii) como parte de trabajos sectoriales independientes y (iv) como productos ex post incluidos en el ciclo regular de evaluación de proyectos del Banco.



Desde mediados de los años 2000, el Banco ha participado en la elaboración de numerosas evaluaciones de impacto. Entre 2006 y 2016, se propuso al menos una evaluación de impacto en más de 400 proyectos.

© BID

1 Antecedentes

El Banco Interamericano de Desarrollo (el BID o el Banco) tiene cada vez más experiencia con la elaboración de evaluaciones de impacto (véase la definición de evaluación de impacto en el Recuadro 1.1), y en el presente informe se hace un balance de lo que se ha hecho hasta la fecha. Desde mediados de los años 2000, el Banco ha participado en la elaboración de numerosas evaluaciones de impacto. Entre 2006 y 2016, se propuso al menos una evaluación de impacto en más de 400 proyectos (de préstamo o de cooperación técnica) y se llegaron a completar casi 100 de ellas. Es el momento de reflexionar sobre lo que este esfuerzo ha reportado para el Banco, el costo que ha acarreado y el rumbo que el Banco deba asumir en adelante. El análisis consta de dos partes, a saber: (1) la evaluación del proceso de producción de las evaluaciones de impacto, en que se comprueba lo que se ha propuesto y la situación correspondiente, el proceso de producción, el costo de las evaluaciones propuestas y la calidad de las que se completaron y de las propuestas; (2) la evaluación del uso y la influencia de las evaluaciones de impacto, en que se consideran el uso interno dado para definir la programación y el diseño de los proyectos del BID, el uso que los clientes del Banco y el público en general han dado a esas evaluaciones, así como el posible efecto que tienen en la implementación de los proyectos.

Las evaluaciones de impacto se han utilizado desde hace mucho tiempo para calcular el efecto causal de una intervención, particularmente en la literatura médica. Desde principios de los años 2000, también se han utilizado ampliamente para evaluar los

efectos de las políticas públicas¹. El propósito de una evaluación de impacto —sobre todo cuando se realiza mediante un ensayo aleatorio controlado— es aislar y medir el efecto de una intervención. Las evaluaciones de impacto generan un mayor entendimiento sobre qué políticas y programas funcionan y cuáles no. Gracias a avances metodológicos y la mayor disponibilidad de datos de alta calidad, sobre todo en el último decenio, el tipo de preguntas que pueden responderse con las evaluaciones de impacto ha aumentado. En un estudio reciente sobre dichas evaluaciones, Dhaliwal y Tulloch (2012) sostienen que las pruebas empíricas que generan las evaluaciones de impacto, cuando se obtienen de un análisis independiente y riguroso desde el punto de vista metodológico, pueden influir en la formulación de políticas y proporcionar pruebas que son fáciles de comprender y llevar a la práctica. Asimismo, las evaluaciones de impacto pueden servir de base para ampliar la escala de políticas públicas y asegurar la sostenibilidad de proyectos de eficacia demostrada.

Los organismos de desarrollo están en una situación privilegiada para fomentar una mejor comprensión de las políticas y los programas, lo que mejora su propia capacidad para patrocinar intervenciones de desarrollo eficaces. Hace más de 10 años, un informe de Center for Global Development (CGD) formuló una pregunta provocadora: ¿Cuándo terminaremos por aprender?² e instó a los organismos de desarrollo a que emplearan evaluaciones de impacto. El concepto en que se sustentó ese informe no es que todos los proyectos de desarrollo deben incluir una evaluación de impacto sino, más bien, que son numerosas las oportunidades para aprender de los proyectos que están ejecutando los organismos internacionales. Asimismo, el informe planteó que, dado el carácter de bien público de conocimiento, la coordinación de todos los organismos en ese sentido reforzaría el proceso de aprendizaje. Alentados por esta agenda basada en pruebas empíricas, los organismos de desarrollo han llevado a cabo muchas iniciativas para reforzar su capacidad de evaluación y recursos, incluidos inversiones para financiar evaluaciones ex post y ensayos aleatorios controlados, lo que ha contribuido enormemente al aumento en la producción de evaluaciones de impacto (Recuadro 1.2).

Aunque las evaluaciones de impacto se utilizan ampliamente para medir la eficacia de la ayuda externa, su utilidad —sobre todo en el caso de los ensayos aleatorios controlados— se ha cuestionado en los últimos años. Las evaluaciones de impacto suscitan cuatro críticas principales.

- Deaton (2009) sostiene que los experimentos no tienen, más que otros métodos, ninguna capacidad especial para producir conocimientos más creíbles, y los experimentos en sí suelen presentar problemas prácticos que socavan toda pretensión de superioridad estadística o epistémica. Independientemente del método utilizado, el autor sostiene que lo importante es tener una teoría clara de los mecanismos que sustentan la intervención objeto de evaluación que, en última instancia, es lo que debe evaluarse.

Recuadro 1.1. ¿Qué es una evaluación de impacto?

Según la OCDE, una evaluación de impacto es una evaluación de cómo la intervención objeto de evaluación afecta los resultados, independientemente de que esos efectos sean intencionales o no (OCDE, 2006). El análisis exige una hipótesis alternativa acerca de los resultados que se habrían obtenido en ausencia de la intervención. Gertler et. al. (2016) describen las evaluaciones de impacto como un tipo particular de evaluación que procura responder a una pregunta específica de causa y efecto: ¿Cuál es el efecto (o efecto causal) de un programa sobre un resultado en particular? El concepto central es diferenciar el efecto real que puede atribuirse a cualquiera de tres casos diferentes: el programa, su modalidad o las innovaciones en su diseño.

Aunque son numerosos los métodos que pueden utilizarse para abordar el tema de causa y efecto, en cada caso la estimación del efecto causal debe basarse en una hipótesis alternativa apropiada: la situación del resultado de interés en ausencia de la política o la intervención en un grupo de características similares. La metodología que se seleccione para la evaluación de impacto debe guardar estrecha relación con las características operativas de la intervención, las reglas de su implementación y la forma en que se recopilan los datos. Cuando estos tres elementos son claros, medir el efecto causal es relativamente fácil [NONIE (2009) y Gertler et. al. (2016)]. A continuación se definen algunos términos pertinentes:

- ***Regresión multivariante.*** Las personas que reciben el tratamiento se comparan con los que no lo reciben, y se neutralizan otros factores que podrían explicar las diferencias en los resultados.
- ***Diferencias en diferencias.*** Se mide la mejora (variación) a lo largo del tiempo que experimentan los participantes del programa, en relación con la mejora (variación) de los que no participan.
- ***Emparejamiento estadístico.*** Las personas de un grupo de control se comparan con un grupo semejante de personas en un grupo experimental.
- ***Diseño de regresión discontinua.*** Las personas se califican sobre la base de criterios específicos y cuantificables. Se fija un valor límite que determina si un individuo reúne las condiciones para participar. Luego, los participantes se comparan con los no participantes y se neutraliza el efecto del criterio de elegibilidad.
- ***Variables instrumentales.*** La participación se predice sobre la base de un factor incidental (casi aleatorio) o una variable “instrumental” que no está correlacionada con el resultado, más allá del hecho de que permite predecir la participación (y la participación afecta el resultado).
- ***Evaluación experimental o ensayo aleatorio controlado.*** Los participantes se asignan aleatoriamente a grupos experimentales o de control para medir la relación causal entre dos variables.

- Muchos investigadores han señalado que una evaluación de impacto por sí sola no es suficiente para orientar las políticas³. Sugieren crear conocimiento utilizando otros tipos de estudios, además de las evaluaciones de impacto. En muchos casos, dichas evaluaciones pueden, por sí solas, responder a interrogantes sobre lo que funciona, pero con frecuencia no abordan la pregunta pertinente de por qué funciona. Cuando se complementan con un seguimiento de procesos y evaluaciones cualitativas, las evaluaciones de impacto pueden ser mucho más informativas, sobre todo para modificar políticas.
- La tercera consideración que despierta preocupación tiene que ver con la validez externa. Muchos investigadores sostienen que los resultados de las evaluaciones de impacto dependen en gran medida del contexto en que se realizan, lo cual limita su capacidad para ser informativas en otros contextos (Deaton, 2010; Pritchett y Sandefur, 2013). Tras analizar más de 20 diversas intervenciones de una serie de estudios que presentaban una gran heterogeneidad de resultados, Vivalt (2015) sostiene que es posible generalizar buscando las diferencias entre las intervenciones y poblaciones seleccionadas. Indica que trabajar de esta forma ayudaría a mejorar la predictibilidad de los resultados de las evaluaciones de impacto.
- Con respecto al debate sobre la validez externa, una inquietud más reciente sobre las evaluaciones de impacto es la atención en los efectos promedio de un programa o política, sin analizar sus efectos distributivos⁴. Muchos investigadores han formulado metodologías diferentes para medir esos efectos.

En resumen, la tendencia del pensamiento académico y la práctica diaria ha sido la de centrar una mayor atención en comprender los mecanismos detrás de las intervenciones en vez de simplemente cuantificar sus resultados, así como en combinar las evaluaciones de impacto con otros tipos de evaluaciones para crear un conocimiento integral sobre lo que funciona.

El Banco ha estado a la vanguardia en la esfera de la producción y el uso de evaluaciones de impacto en América Latina y el Caribe, sobre todo desde que participó en la evaluación de PROGRESA en México. El BID proporcionó financiamiento para la evaluación de impacto de PROGRESA a cargo del Instituto Internacional de Investigación sobre Políticas Alimentarias y siguió estrechamente su evolución⁵. Pocos años más tarde adoptó como práctica óptima el uso de evaluaciones de impacto en proyectos similares⁶. Este esfuerzo inicial del BID se concentró en los sectores sociales debido a la disponibilidad de datos y la familiaridad con los métodos de evaluación de muchos países. En otros sectores, el proceso de crear una cultura de evaluación basada en evaluaciones de impacto ha sido más lento.

A partir de principios de los años 2000, OVE empezó a presionar al BID para que se concentrara más en la evaluación. En 2002, OVE publicó su primer análisis de la evaluabilidad del Banco, estableció las normas básicas para que un proyecto pueda evaluarse y demostró que en la mayoría de los proyectos el BID estaba desaprovechando

la oportunidad para profundizar los conocimientos sobre eficacia⁷. En el Informe de Evaluabilidad que OVE publicó en 2005 se recomendaba la adopción de normas de evaluabilidad como criterio de aprobación de proyectos. La recomendación se adoptó y se incluyó en el mandato del Banco como parte del Noveno Aumento General de Recursos. Mientras OVE preconizaba la inclusión de un mayor grado de evaluabilidad en la etapa de diseño de los proyectos, también empezó a formular su propia agenda de evaluaciones de impacto ex post. En 2004, el Directorio Ejecutivo le encomendó a OVE el mandato de realizar evaluaciones de impacto de forma sistemática de los proyectos del BID⁸.

Recuadro 1.2. Las evaluaciones de impacto en otros organismos

Con la creación de la iniciativa Evaluación del Impacto sobre el Desarrollo (DIME) en 2005, el Banco Mundial empezó a emplear las evaluaciones de impacto en forma más sistemática para medir los resultados, promover el aprendizaje y orientar las estrategias institucionales. La iniciativa contribuyó a incrementar el número de evaluaciones de impacto completadas en el Banco Mundial, de 28 en 2005 a 215 en 2015. A nivel de los programas, DIME facilita la formulación del programa de investigación, la gestión de los esfuerzos de investigación y la realización de las actividades de divulgación de información, difusión de políticas y recaudación de fondos.

Millennium Challenge Corporation (MCC) también ha destinado fondos a las evaluaciones de impacto y sus mandatos fueron objeto de una revisión para exigir dichas evaluaciones. Creado por el Congreso de los Estados Unidos en 2004, Millennium Challenge Corporation ha procurado prestar asistencia externa centrando la atención en políticas de buena calidad, la identificación de los países con los programas y resultados. En los seis primeros años dio un fuerte impulso a las evaluaciones de impacto como parte de su marco de resultados. En 2011, más del 60% de las evaluaciones de sus proyectos eran evaluaciones de impacto. Actualmente, la organización propone esas evaluaciones siguiendo un enfoque más estratégico, basándose en un marco de factibilidad revisado. No obstante, en 2016 casi el 40% de los proyectos incluyeron evaluaciones de impacto.

Asimismo, la comunidad académica ha establecido redes que respaldan el uso de evaluaciones de impacto para estudiar las políticas de desarrollo. Entre las más destacadas figura Abdul Latif Jameel Poverty Action Lab (J-PAL), creada en 2003 en el Departamento de Economía del Instituto de Tecnología de Massachusetts (MIT), y colabora con más de 20 gobiernos de 11 países para institucionalizar el uso de pruebas empíricas en la toma de decisiones. Desde su creación ha emprendido 842 evaluaciones de impacto, en marcha y completadas, a través de una red de más de 140 catedráticos afiliados en 49 universidades de todo el mundo. La organización cuenta con oficinas regionales independientes auspiciadas por importantes universidades de África, Europa, América Latina y el Caribe, América del Norte, Asia Meridional y Asia Sudoriental.

Fue solo después de la realineación y el comienzo de la agenda del Banco de Conocimiento que se crearon las condiciones para promover las evaluaciones de impacto a cargo de las unidades operativas del BID. Dos cambios fundamentales aceleraron ese proceso. El

primero fue el cambio de perfil del personal, pues con la realineación el Banco pudo ampliar su capacidad interna de investigación, incluidas las evaluaciones de impacto, y aumentó sustancialmente el número de especialistas de mayor formación académica⁹. El segundo fue la creación de la Oficina de Planificación Estratégica y Efectividad en el Desarrollo (SPD). Conforme señala el documento sobre la realineación, “la creación de la Oficina de Planificación Estratégica y Efectividad en el Desarrollo es una señal inequívoca del cambio cultural que persigue la realineación”. Parte del mandato de SPD era ayudar a asegurar la efectividad del Banco en el desarrollo “proporcionando conocimiento técnico en el ciclo de proyectos”¹⁰.

Al favorecer el uso de evaluaciones de impacto, la intención del Banco ha sido fomentar intervenciones más eficaces. El Marco de Efectividad en el Desarrollo¹¹, preparado por SPD y aprobado en 2008, internalizó la idea de que el Banco mejoraría su efectividad en el desarrollo a través de la comprensión cabal de las intervenciones que funcionan y las que no funcionan. El marco permitió definir mejor las funciones y responsabilidades de las unidades del Banco en el ámbito de la efectividad en el desarrollo e introdujo la DEM para determinar la evaluabilidad ex ante de las operaciones. La DEM creó una serie de requisitos de seguimiento y evaluación que sirven de base para calificar las operaciones en función de su cumplimiento ex ante. El sistema reconocía explícitamente la importancia de estimar la eficacia causal de las operaciones al asignar mayores calificaciones a las que incluían evaluaciones de impacto en su plan de evaluación. Así pues, desde 2009 la DEM ha incentivado claramente la elaboración de evaluaciones de impacto.

El cambio cultural hacia la generación de intervenciones basadas en pruebas empíricas también fue un componente fundamental del Noveno Aumento y la nueva Estrategia Institucional del BID. Ambos documentos tuvieron por objetivo afianzar el marco institucional para generar productos de conocimiento que se ajustasen más a las necesidades de la región. Según la Actualización de la Estrategia Institucional 2010 2020:

“Para promover la adaptación de enfoques eficaces de desarrollo, es importante fortalecer la capacidad del Banco para aprender y difundir lo que ha aprendido. Esto exige incrementar los esfuerzos desplegados para generar conocimientos sobre qué funciona y qué no funciona con base en información rigurosa, lo que, a su vez, requiere una continua evaluación de los proyectos durante su ciclo de vida. Por otra parte, para que la innovación pueda impregnar a toda la institución, el BID necesita crear una nueva cultura en que la colaboración, la diversidad, la experimentación y la práctica sean los principales ingredientes que resulten en mejores soluciones, procesos y productos”¹².

En este contexto, OVE evaluó la producción y el uso de las evaluaciones de impacto en las operaciones del sector público del BID¹³. El análisis se centró en el sector público porque, hasta la fecha, casi todos los esfuerzos emprendidos han sido en ese sector. OVE recopiló datos detallados sobre todas las operaciones de préstamo y de cooperación técnica aprobadas entre 2006 y 2016 que incluían al menos una evaluación de impacto (Recuadro 2.1). En vista de que el Marco de Efectividad en el Desarrollo fue aprobado

en 2008 y la DEM se implementó en 2009, OVE decidió incluir varios años previos a esos como referencia. Asimismo, OVE complementó su análisis de la producción de evaluaciones de impacto en el Banco contratando los servicios del Centro para el Aprendizaje en Evaluación y Resultados para América Latina y el Caribe (CLEAR-LAC) (mediante un proceso de licitación) para que examinara sistemáticamente la calidad de la mayoría de las evaluaciones de impacto completadas y de una muestra de propuestas de evaluaciones de impacto en marcha. Además, OVE llevó a cabo un análisis de las citas a evaluaciones de impacto publicadas y entrevistó a una muestra de clientes del Banco para recabar información sobre las opiniones de los clientes sobre las evaluaciones de impacto y su uso. En total, OVE realizó más de 200 entrevistas a empleados del Banco que usan o elaboran evaluaciones de impacto, gerentes sectoriales, jefes de división, asesores económicos regionales y representantes de otros organismos que trabajan con dichas evaluaciones. En el Cuadro 1.1 figura un resumen de los métodos empleados.

CUADRO 1.1 - PREGUNTAS DE EVALUACIÓN Y MEDIOS DE VERIFICACIÓN

Tema	Pregunta de evaluación	Medio utilizado para abordar la pregunta
Análisis de la producción	¿Cuántas evaluaciones de impacto se han propuesto?	<ul style="list-style-type: none"> Creación de una base de datos de todas las evaluaciones de impacto propuestas en los documentos de operaciones de préstamo y cooperación técnica, sobre la base de un análisis documental y entrevistas con personal encargado del Banco.
	¿En qué situación se encuentran las evaluaciones de impacto propuestas?	
	¿Cuál fue el proceso de originación, diseño e implementación?	
	¿Cuánto cuestan las evaluaciones de impacto?	<ul style="list-style-type: none"> Entrevistas a especialistas, jefes de división, gerentes sectoriales y clientes del Banco.
	¿Cuáles son las prácticas óptimas en materia de financiamiento de las evaluaciones de impacto y los productos de conocimiento en general?	<ul style="list-style-type: none"> Entrevistas a empleados del Banco que participan en las evaluaciones de impacto y representantes de otras organizaciones que trabajan con evaluaciones de impacto.
	¿Cuál es la calidad de las evaluaciones de impacto del BID y cómo ha evolucionado a lo largo del tiempo?	<ul style="list-style-type: none"> Evaluación de CLEAR basada en criterios estándar de 84 evaluaciones completadas y 59 propuestas de evaluaciones en marcha.
Análisis del uso y grado de influencia	¿Cómo se emplea en el Banco el conocimiento generado por las evaluaciones de impacto?	<ul style="list-style-type: none"> Análisis documental de estrategias de país, notas técnicas y documentos de marco sectorial. Análisis de citas en préstamos aprobados entre 2007 y 2016. Entrevistas a jefes de división, gerentes sectoriales y asesores económicos regionales. Entrevistas a empleados del Banco que trabajan con evaluaciones de impacto.
	¿Cómo emplean los clientes el conocimiento generado por las evaluaciones de impacto?	<ul style="list-style-type: none"> Entrevista a una muestra de 25 clientes del Banco relacionados con proyectos que incluían evaluaciones de impacto.
	¿Cómo se emplean las evaluaciones de impacto en general?	<ul style="list-style-type: none"> Análisis de citas a todas las evaluaciones de impacto publicadas entre 2006 y 2016.
	¿Facilitan las evaluaciones de impacto la ejecución de los proyectos?	<ul style="list-style-type: none"> Análisis econométrico de los desembolsos de los proyectos.

En el presente documento, OVE se concentra en las evaluaciones de impacto propuestas en documentos de operaciones de préstamo y cooperación técnica.

© BID

2 La Producción de Evaluaciones de Impacto en el BID

El Banco no mantiene una base de datos centralizada, un registro oficial o una clasificación presupuestaria de las evaluaciones de impacto y, por lo tanto, no existe un mecanismo institucional que permita determinar fácilmente su producción. Asimismo, como las evaluaciones de impacto pueden elaborarse en muchos contextos diferentes, es difícil identificarlas en función de actividades o registros¹⁴. En el presente documento, OVE se concentra en las evaluaciones de impacto propuestas en documentos de operaciones de préstamo y cooperación técnica. En el Recuadro 2.1 se resume la metodología utilizada para identificar el universo de evaluaciones de impacto y sus limitaciones (véase el Anexo I).

A. ANÁLISIS DE LA PRODUCCIÓN

1. ¿Cuántas evaluaciones de impacto se han propuesto y en qué situación se encuentran?

Los cambios institucionales que se llevaron a cabo en el BID, sobre todo la adopción de la DEM, dieron lugar a un marcado aumento del número de operaciones de préstamo y cooperación técnica con evaluaciones de impacto. Entre 2006 y 2008, un promedio de 8% de las operaciones de préstamo y 2% de las de cooperación técnica incluyeron una evaluación de impacto. Estas cifras aumentaron a un promedio de 28% y 3%, respectivamente, en el período comprendido entre 2009 y 2016. Entre 2006 y 2008, se propusieron en promedio 13 evaluaciones de impacto al año, frente a un promedio anual de 61 evaluaciones de impacto entre 2009 y 2016. El número de propuestas aumentó notablemente tras la adopción de la DEM hasta 2013 y luego disminuyó ligeramente (Gráficos 2.1 y 2.2).

Recuadro 2.1. Metodología

Debido a la dispersión de los documentos y la multiplicidad de fuentes, OVE empleó una estrategia de cuatro pasos para maximizar las probabilidades de reconocer exitosamente cada evaluación de impacto realizada o financiada por el Banco.

- Exploración inicial de documentos de préstamo y perfiles de operaciones de cooperación técnica. Análisis léxico de todos los documentos de préstamo registrados en IDB-SEC y de la versión más reciente del perfil de las operaciones de cooperación técnica publicadas en IDBDOCS entre 2006 y 2016, seguido por un análisis documental integral de dichos documentos.
- Entrevistas a jefes de equipo y jefes de división. Entrevistas estructuradas con los jefes de equipo o el especialista encargado de la evaluación de impacto en cada división para recabar datos no incluidos en las propuestas de préstamo, los anexos de seguimiento y evaluación o los perfiles de las operaciones de cooperación técnica.
- Proceso de validación con SPD. Una comparación de la base de datos de OVE con todas las evaluaciones de impacto registradas en la base de datos de SPD.
- Procesamiento de base de datos. La creación de una base de datos con toda la información pertinente para cada evaluación de impacto identificada. El análisis del presente informe se sustenta en esa base de datos.

OVE identificó 414 operaciones (122 operaciones de cooperación técnica y 292 operaciones de préstamo) que incluyeron al menos una evaluación de impacto. Puesto que muchos préstamos y operaciones de cooperación técnica incluyeron múltiples evaluaciones de impacto, la base de datos de OVE cuenta con información sobre 531 evaluaciones de impacto propuestas entre 2006 y 2016. Algunas de estas evaluaciones se diseñaron y se llevarán a cabo después de aprobarse las operaciones y, en esos casos, OVE no contó con un mecanismo para realizar el seguimiento.

OVE llevó a cabo un trabajo complementario para identificar la producción de evaluaciones de impacto en el BID y compararla con las tendencias globales. Puesto que las evaluaciones de impacto también pueden prepararse utilizando recursos de actividades económicas y sectoriales y del presupuesto administrativo, el enfoque de OVE no identificó todas las evaluaciones de impacto del Banco completadas o en marcha. Por ese motivo, creó una base de datos secundaria para las evaluaciones de impacto publicadas que incluye información sobre estudios individuales (título, año, autor y departamento) y enlaces a esos estudios. El número de publicaciones del BID basadas en evaluaciones de impacto ha aumentado marcadamente desde 2009 y su tendencia parece ajustarse a la que describen Cameron, Mishra y Brown (2016) en cuanto a su crecimiento, distribución sectorial y metodologías. Aunque en un inicio las evaluaciones de impacto se concentraron sobre todo en los sectores sociales, con el tiempo otros sectores han elaborado un mayor número de evaluaciones de impacto.

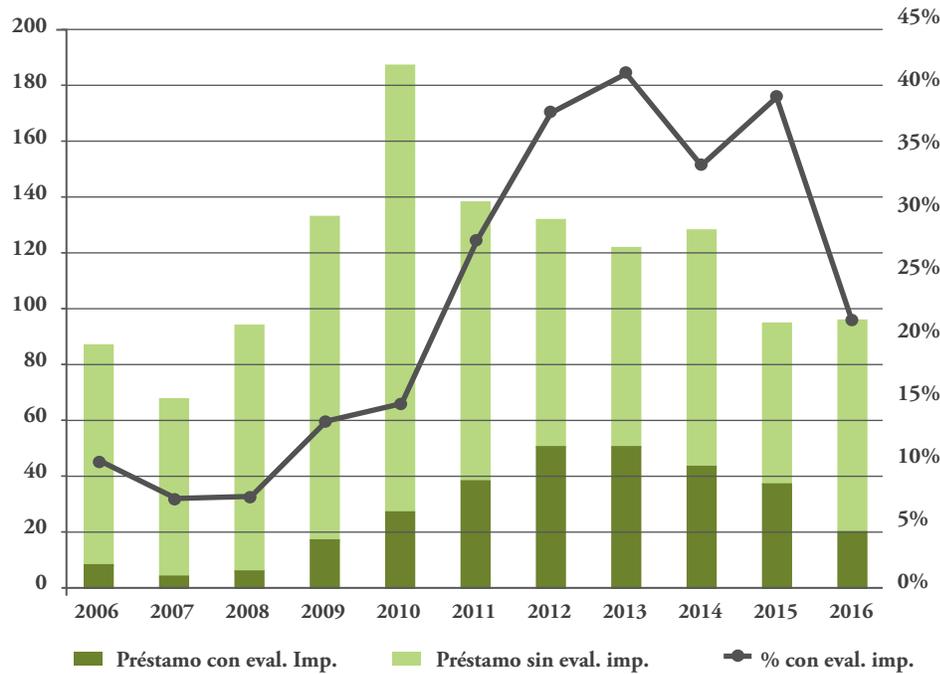


GRÁFICO 2.1
Operaciones de préstamo del BID con evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

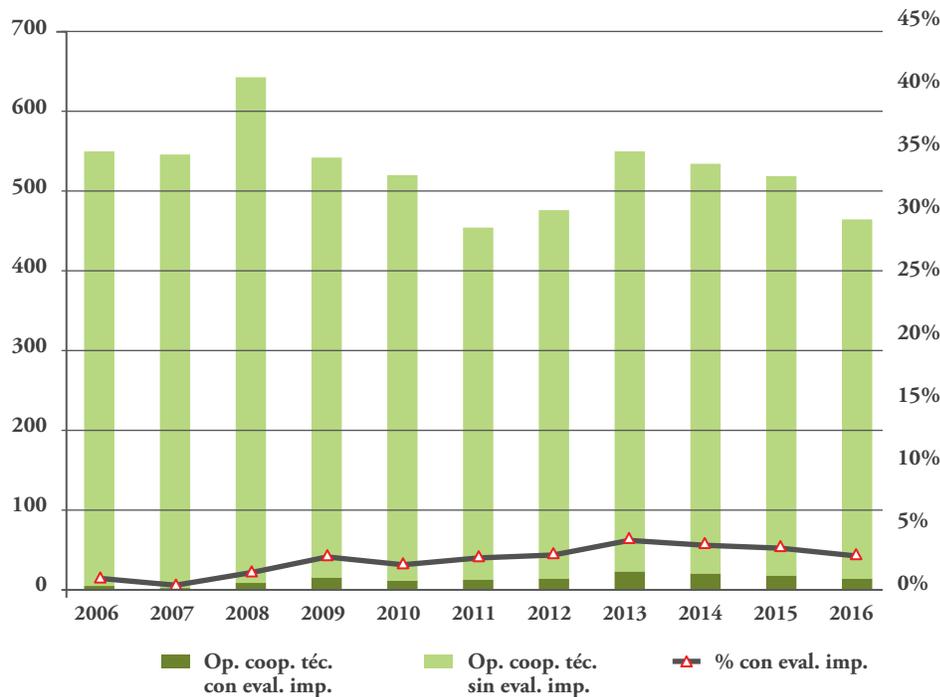


GRÁFICO 2.2
Operaciones de cooperación técnica del BID con evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

Se incluyeron un total de 531 evaluaciones de impacto en las operaciones de préstamo y de cooperación técnica aprobadas entre 2006 y 2016. De las evaluaciones de impacto propuestas, 94 (18%) se han completado, 151 (28%) se han cancelado¹⁵ y 286 (54%) siguen en el inventario. Conforme se refleja en el Cuadro 2.1, SCL ha propuesto la

mayoría de las evaluaciones de impacto (47%), seguido por IFD (28%) y CSD (14%). INT ha completado 15 evaluaciones de impacto con financiamiento del programa de actividades económicas y sectoriales, que no se reflejan en el Cuadro 2.1. La situación de las evaluaciones de impacto en curso varía de la fase de diseño (40%) a la de análisis final de los datos (11%), y se prevé que la mayoría se habrá completado dentro de los próximos cinco años¹⁶.

CUADRO 2.1. DISTRIBUCIÓN DE LAS EVALUACIONES DE IMPACTO, POR DEPARTAMENTO, DIVISIÓN Y SITUACIÓN

Departamento	División	Operaciones de préstamo				Operaciones de cooperación técnica			
		Completadas	En marcha	Canceladas	Total	Completadas	En marcha	Canceladas	Total
CSD	CCS	0	1	1	2	0	1	0	1
	RND	6	33	9	48	1	2	2	5
	HUD	1	9	9	19	0	1	0	1
IFD	CMF	3	8	10	21	0	0	3	3
	CTI	5	26	8	39	2	3	3	8
	FMM	0	5	2	7	0	0	0	0
	ICS	3	29	12	44	7	13	8	28
INE	ENE	0	3	1	4	0	0	0	0
	TSP	1	12	0	13	0	1	0	1
	WSA	0	5	3	8	0	1	1	2
INT	TIN	0	20	2	22	0	0	1	1
SCL	EDU	9	19	21	49	17	11	6	34
	GDI	1	2	0	3	6	7	5	18
	LMK	1	5	7	13	11	3	4	18
	SPH	11	58	25	94	9	8	8	25
Total		41	235	110	386	53	51	41	145

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017; OVEDA (última columna); incluye evaluaciones de impacto realizadas por VPC (CAN y CCB), RES y la dirección de SCL, clasificadas por tema y sector.

Las entrevistas que realizó OVE a jefes de equipo¹⁷ indican que las principales razones por las que se cancelaron evaluaciones de impacto, cuando no se cancelaron en sí las operaciones de préstamo o de cooperación técnica, parecen haber sido desafíos políticos (30%) y temas relacionados con la implementación y el diseño (34%) (Gráfico 2.3). Las evaluaciones de impacto que se propusieron principalmente para aumentar el puntaje en la DEM, en vez de como resultado de un interés real del BID o el cliente en aprender, tenían una probabilidad mucho mayor de que fueran canceladas¹⁸. La adopción de un método bien definido para las evaluaciones de impacto se vincula a una menor probabilidad de cancelación, sobre todo cuando el método es el del ensayo aleatorio controlado. De hecho, si se inicia un ensayo de ese tipo que luego fracasa por cualquier motivo, es más probable que los equipos encuentren otra forma de realizar una evaluación de impacto¹⁹. La probabilidad de cancelación no varía considerablemente por país o por el costo de la evaluación de impacto.

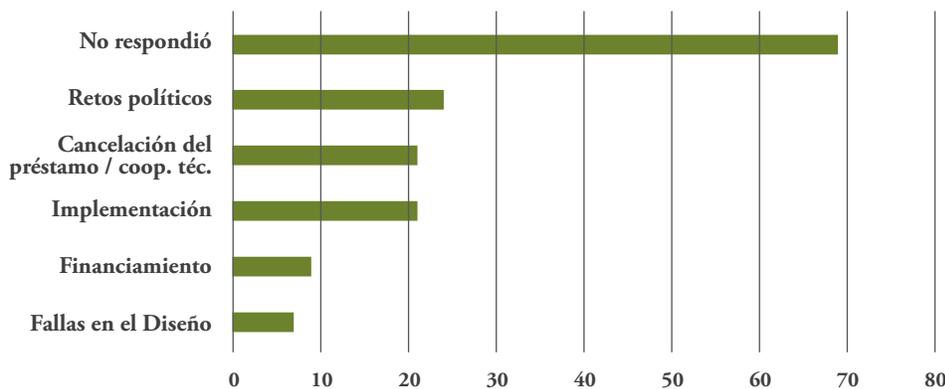


GRÁFICO 2.3

Principales razones de cancelación de las evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

La distribución de las evaluaciones por departamentos de países y sectores ha sido desigual. En cuanto a la distribución por país, CID tiene el porcentaje más elevado (35%) y CCB el más bajo (7%) (véase el Anexo II). En lo que respecta a la distribución por sector, las evaluaciones de impacto completadas se concentran enormemente en proyectos anteriores (Gráfico 2.4) en los sectores sociales (Gráfico 2.5). En otros sectores se han completado pocas evaluaciones de impacto, si bien el número (particularmente en IFD) ha ido en aumento²⁰. Algunas evaluaciones se han financiado mediante actividades económicas y sectoriales (en especial en INT) y no se reflejan aquí²¹. Los sectores sociales son los que tradicionalmente han realizado más evaluaciones de impacto debido en parte a que en esos sectores es más fácil observar quiénes son los beneficiarios y establecer una clara metodología para evaluar el efecto de una intervención²². No obstante, sectores en los que es extremadamente difícil realizar evaluaciones de impacto han emprendido cada vez mayores esfuerzos para innovar utilizando diferentes metodologías²³. Además, OVE observó algunas diferencias en las razones por las que se cancelaron evaluaciones de impacto, siendo los principales motivos en SCL, INE e INT factores políticos y temas de implementación, en tanto que en CSD e IFD los motivos fueron temas de diseño y la cancelación de operaciones subyacentes.

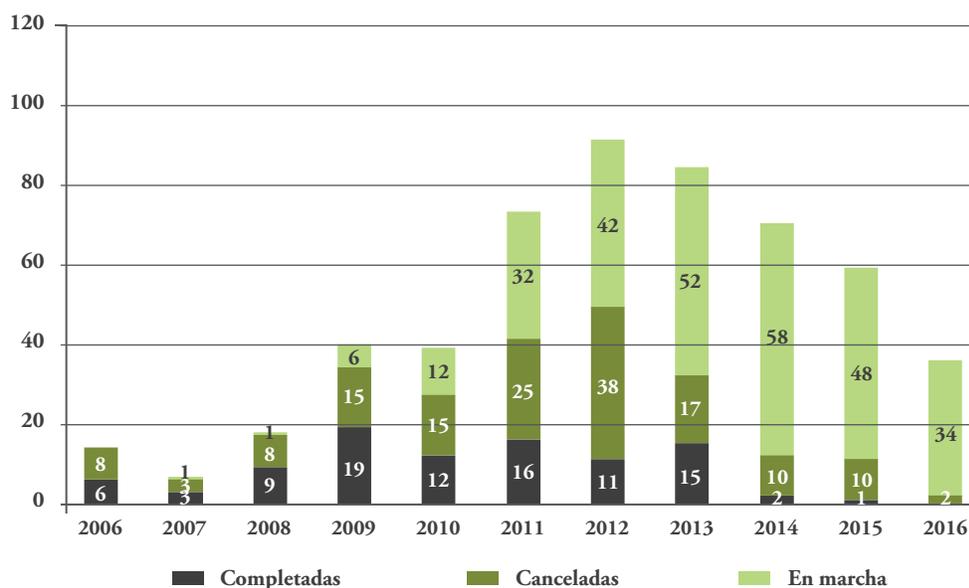


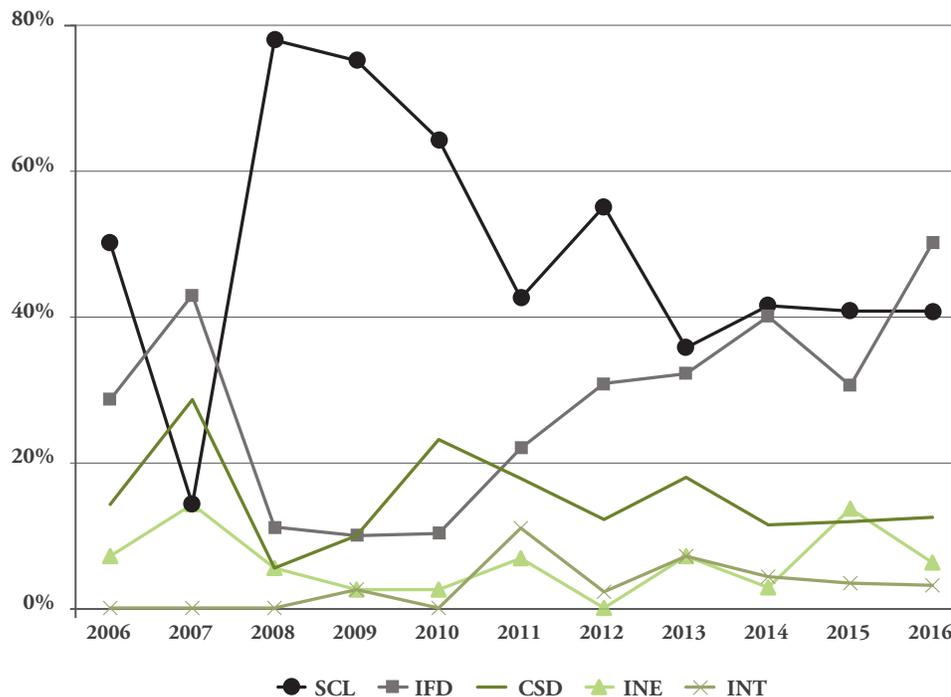
GRÁFICO 2.4

Situación de las evaluaciones de impacto, por año de aprobación

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

GRÁFICO 2.5
Distribución de las evaluaciones de impacto, por departamento

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



El Banco ha diversificado no solo los sectores que realizan evaluaciones de impacto, sino también el tipo de metodologías que se emplean. El número de evaluaciones propuestas sin una metodología definida ha disminuido (Gráfico 2.6); en 2016 todos los proyectos tenían una metodología clara al momento de aprobación. Ello indica que se está prestando más atención al diseño de las evaluaciones de impacto y al mecanismo causal que se está poniendo a prueba en los proyectos. En la medida en que las evaluaciones que incluyen una metodología bien definida tengan menor probabilidad de cancelarse (como se señaló anteriormente), con el tiempo debería disminuir el número de cancelaciones. El número general de ensayos aleatorios controlados, en valores absolutos y relativos, ha bajado, y el número de evaluaciones de impacto cuasiexperimentales ha aumentado en forma correspondiente. Si bien todavía es necesario recopilar datos, desde 2011 un mayor número de evaluaciones han utilizado datos administrativos.

Las tendencias observadas más arriba tienen algunas repercusiones importantes para el Banco. En primer lugar, muchas evaluaciones han sido canceladas²⁴, lo cual no es, de por sí, un resultado negativo; en algunos casos, realizar otro tipo de evaluación podría proporcionar más información que una evaluación de impacto que no cumpla las condiciones para poder completarse y utilizarse. No obstante, puede surgir un problema importante si la evaluación de impacto se acompaña de deficiencias generales en la planificación de las tareas de seguimiento y evaluación. Si la evaluación de impacto es la única fuente de información y no se cuenta con un sistema de seguimiento y evaluación adecuado, cancelarla significará que el Banco no dispondrá de información sobre los resultados del proyecto²⁵. Segundo, el Banco ha generado una gran cartera

de evaluaciones de impacto en marcha. A medida que estas evaluaciones se vayan completando, será cada vez más importante que el Banco cuente con un sistema que permita hacer un seguimiento de lo que se está produciendo, como se hace mediante BRIK para las publicaciones, y que también permita asimilar el conocimiento generado, con una estrategia para organizar y difundir los hallazgos. Por último, la reducción del número propuesto de evaluaciones de impacto y de evaluaciones sin metodología apunta a que el Banco posiblemente esté empezando a hacer un uso más estratégico de las evaluaciones de impacto.

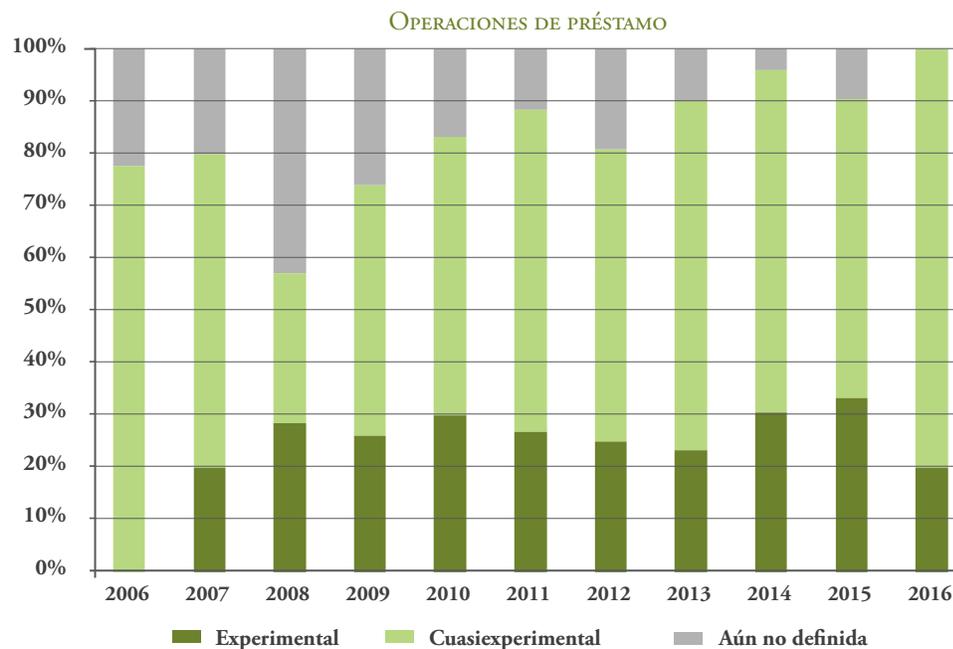


GRÁFICO 2.6
Distribución de las evaluaciones de impacto, por metodología y año

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



2. ¿Cómo han funcionado los procesos de originación, diseño e implementación?

El principal problema planteado por el marcado aumento de la producción de evaluaciones de impacto es la falta de estructura del proceso de generación. El Banco ha establecido grandes incentivos para proponer evaluaciones de impacto, pero hasta la fecha no ha creado un mecanismo para proponerlas en forma estratégica. En 2012, el Banco adoptó un nuevo marco normativo para las intervenciones sectoriales e introdujo un documento de marco sectorial para “proporcionar orientación significativa a los equipos de proyecto y ofrecer una idea clara de lo que el Banco procura lograr en un determinado sector” (documento GN 2670-1). Se supone que los documentos de marco sectorial, que tienen una estructura bien definida que destaca la importancia de las pruebas empíricas y la efectividad en el desarrollo, deben incluir una revisión integral de las pruebas empíricas internacionales sobre los principales retos de la región. Aunque dichos documentos han de identificar la labor clave de conocimiento por sector, la medida en que los sectores logran identificar proactivamente las brechas de conocimiento y traducirlas en líneas de acción específicas varía de un sector a otro. Una minuciosa revisión de los 20 documentos de marco sectorial vigentes (véase el Anexo VII) indica que solo alrededor del 30% presenta una revisión cabal de la literatura especializada en que se identifican brechas de conocimiento y la medida necesaria para orientar la priorización de las evaluaciones de impacto. Más de la mitad de esos documentos de marco sectorial se hicieron en el sector social, lo cual no sorprende dado el grado de desarrollo alcanzado en la producción de evaluaciones de impacto en los sectores de educación, protección social, atención a la salud y género.

Como indica el análisis anterior, los grandes incentivos para producir evaluaciones de impacto parecen haber causado en cierto nivel un número excesivo de propuestas, y muchas evaluaciones de impacto no reúnen los criterios para ser pertinentes o incluso factibles. Realizar una evaluación de impacto no es fácil, dado el tiempo, los recursos y la voluntad política que se necesitan para su éxito²⁶. Es comprensible que algunas de ellas se cancelen durante la implementación del proyecto en cuestión. Sin embargo, como se señaló anteriormente, es más probable que se cancele una evaluación de impacto cuya propuesta haya sido motivada principalmente por la DEM, porque es posible que no haya un verdadero interés ni un compromiso del gobierno y el equipo de proyecto. En las operaciones de préstamo, alrededor del 34% de las evaluaciones de impacto propuestas fueron motivadas únicamente por el puntaje en la DEM (Gráfico 2.7). En el 17% de las cancelaciones, los especialistas informaron que no estaban dadas las condiciones para implementar una evaluación de impacto y, más adelante, la implementación se hizo imposible.

Un gran porcentaje de las evaluaciones de impacto revisadas —el 46% de las incluidas en préstamos y el 66% de las incluidas en operaciones de cooperación técnica— fueron impulsadas por el Banco (Gráficos 2.6 y 2.7). Solo el 14% de

las evaluaciones de impacto incluidas en préstamos y el 20% de las incluidas en operaciones de cooperación técnica pueden atribuirse al interés manifestado por el cliente. Los entrevistados observaron que muchas evaluaciones de impacto no examinaron preguntas pertinentes para el cliente y, con frecuencia, se presentaron al cliente como un requisito del Banco. En consecuencia, los gobiernos muchas veces percibieron las evaluaciones de impacto como una carga, en lugar de un producto o servicio valioso que apoya la implementación de un proyecto.

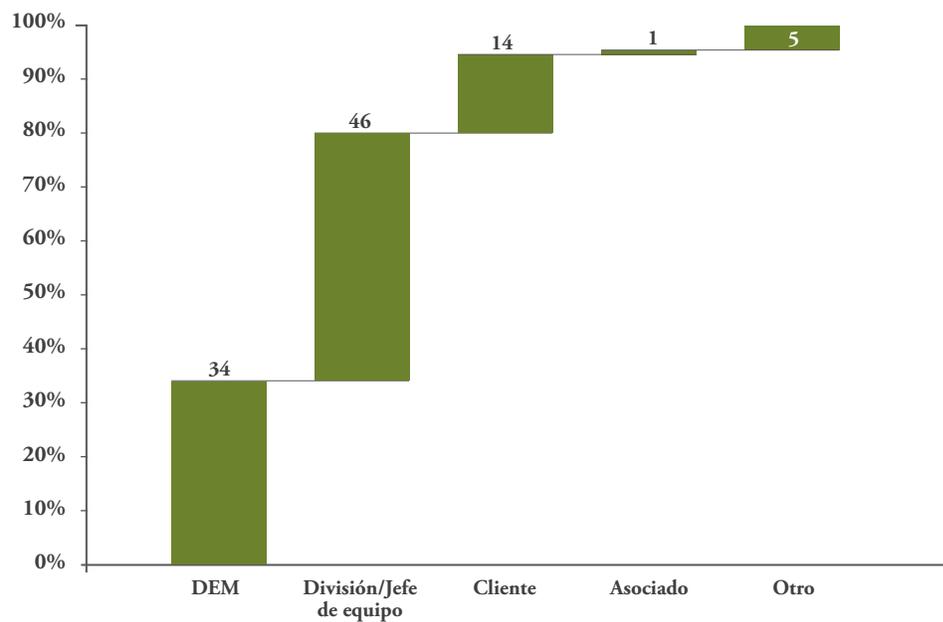


GRÁFICO 2.7
Origen de evaluaciones de impacto en operaciones de préstamo (%)

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

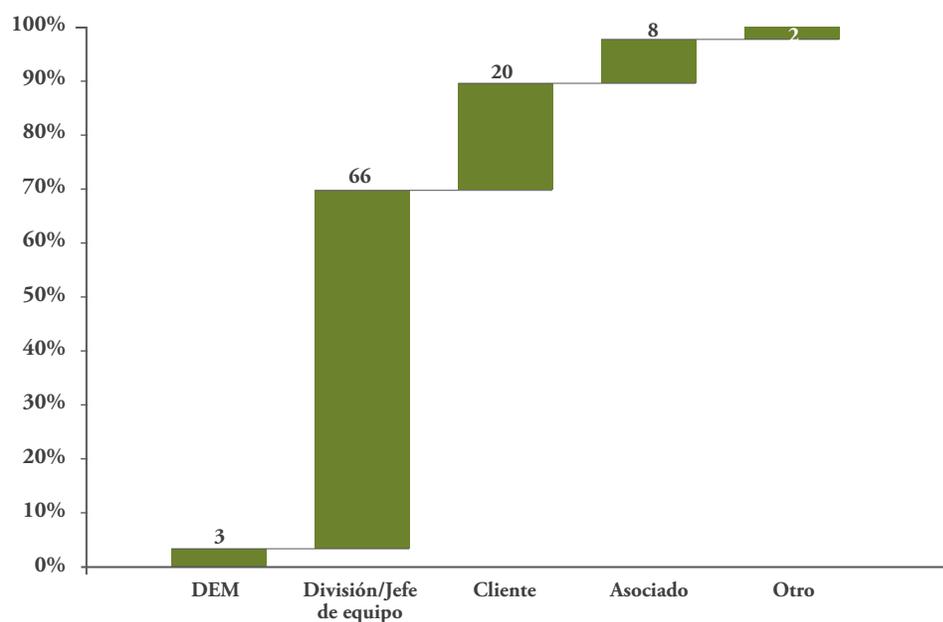


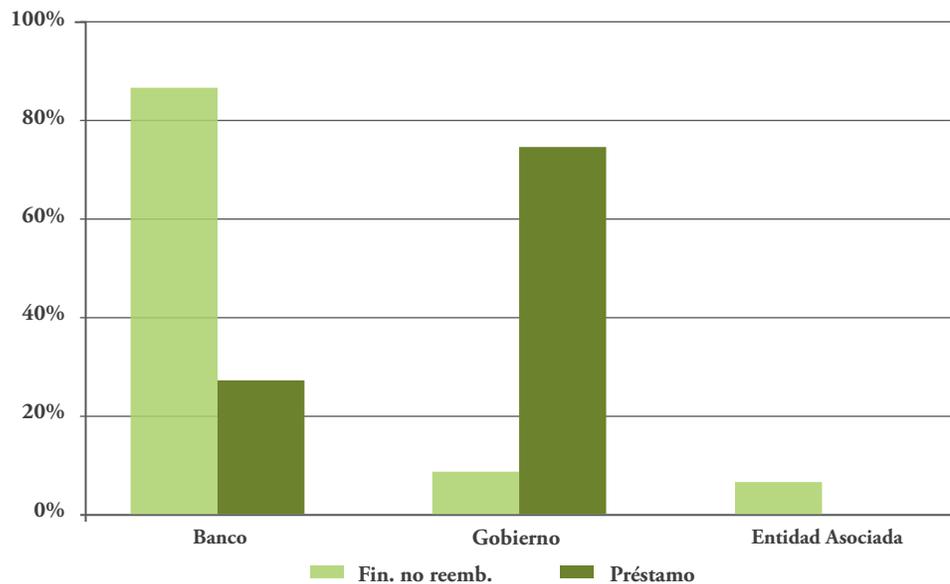
GRÁFICO 2.8
Origen de evaluaciones de impacto en operaciones de cooperación técnica (%)

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

Cuando se preguntó al personal sobre la propiedad de las evaluaciones de impacto, la mayor parte contestó no estar muy seguro. Una visión común fue que las evaluaciones financiadas por préstamos pertenecen al gobierno en cuestión, mientras que en el caso de las operaciones de cooperación técnica, la evaluación pertenece al Banco (Gráfico 2.9), lo cual es comprensible particularmente en el caso de operaciones de cooperación técnica administradas por el Banco. No obstante, en alrededor del 30% de las evaluaciones de impacto incluidas en préstamos y el 10% de las incluidas en operaciones de cooperación técnica, el jefe de equipo no tenía claro a quién pertenecía la evaluación o cómo proceder para consultar a los clientes sobre la divulgación de los resultados. Los clientes tienen la misma confusión: en muchos casos no saben qué hacer con la evaluación más allá de los usos internos que se le pueda dar, incluso en los casos en que los resultados puedan ser pertinentes para otros²⁷.

GRÁFICO 2.9
Percepción del jefe de equipo en cuanto a la propiedad de la evaluación de impacto, por fuente del presupuesto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



El personal del Banco participa principalmente en el diseño de las evaluaciones de impacto. El personal del Banco diseña, en promedio, el 60% de las evaluaciones de impacto, cifra que ascendió al 80% en 2016 (Gráfico 2.10). Ese porcentaje no varía de manera importante entre los diferentes sectores. Ello responde en parte al mayor énfasis del Banco en la calidad del diseño y el reconocimiento de que no todas las entidades públicas cuentan con la misma capacidad para producirlas. Una vez aprobado el proyecto, la realización de la evaluación de impacto pasa a ser responsabilidad del cliente, que a menudo (en el 70% de los casos) contrata a consultores para que la hagan, sobre todo en el caso de los préstamos²⁸. El Banco muchas veces apoya a los gobiernos en la formulación de términos de referencia para contratar consultores y supervisar su labor. Algunos clientes ven ese apoyo como un servicio positivo que ofrece el Banco, aunque otros no están interesados y perciben las evaluaciones de impacto como un requisito más de la institución, por lo que dejan a que el personal lo haga por su cuenta.

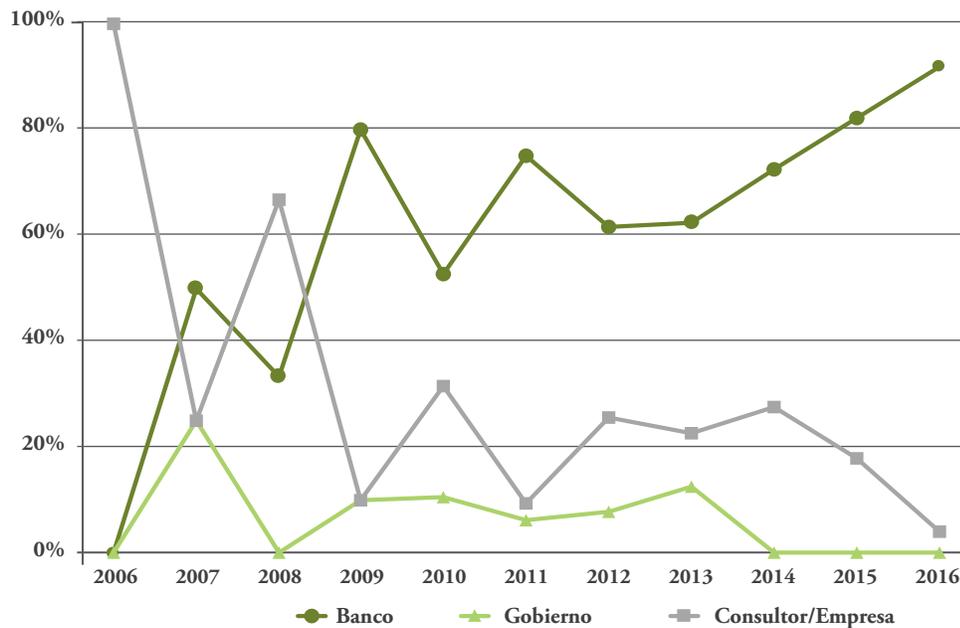


GRÁFICO 2.10
Diseño de evaluaciones de impacto, por año

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

SPD ha tenido un papel fundamental a través del control de calidad de la DEM y de una participación más directa en alrededor del 25% de las evaluaciones de impacto. El personal del Banco indicó que había recibido un importante apoyo de SPD en el 35% y el 46% de las evaluaciones de impacto incluidas en operaciones de cooperación técnica y de préstamo, respectivamente (Gráfico 2.11). El apoyo de SPD se ha concentrado en sectores con menos pericia en evaluaciones de impacto, habiendo recibido INE el mayor nivel de apoyo (Gráfico 2.12). Actualmente, SPD brinda apoyo directo a 170 evaluaciones de impacto en marcha. Más allá de asegurar una calidad mínima a través del proceso de revisión, SPD participa más activamente en el diseño y a veces en la implementación.

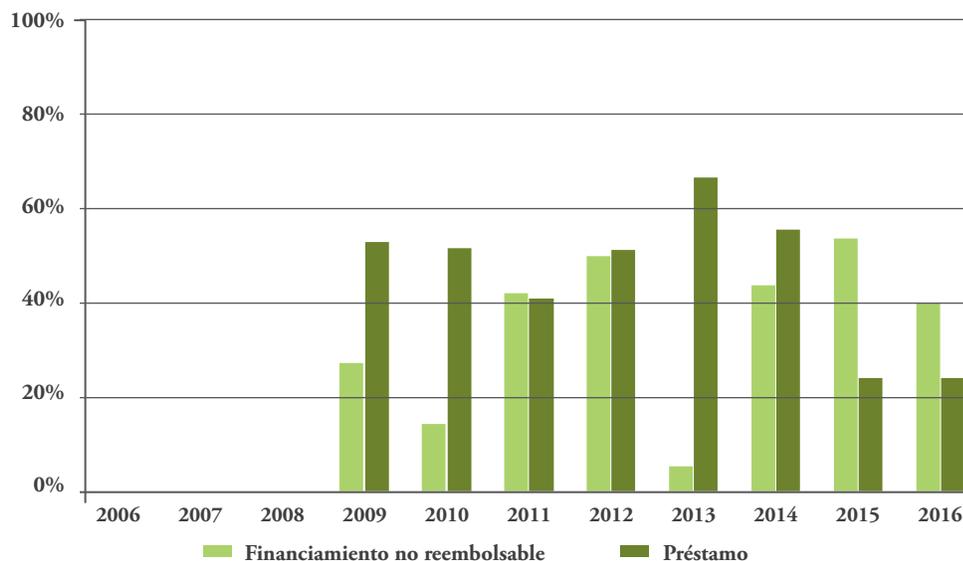


GRÁFICO 2.11
Apoyo de SPD, por año

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

Nota: Las operaciones de financiamiento no reembolsable incluyen operaciones de cooperación técnica y operaciones de financiamiento no reembolsable para inversión.

GRÁFICO 2.12
Porcentaje de evaluaciones de impacto con apoyo de SPD, por sector

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

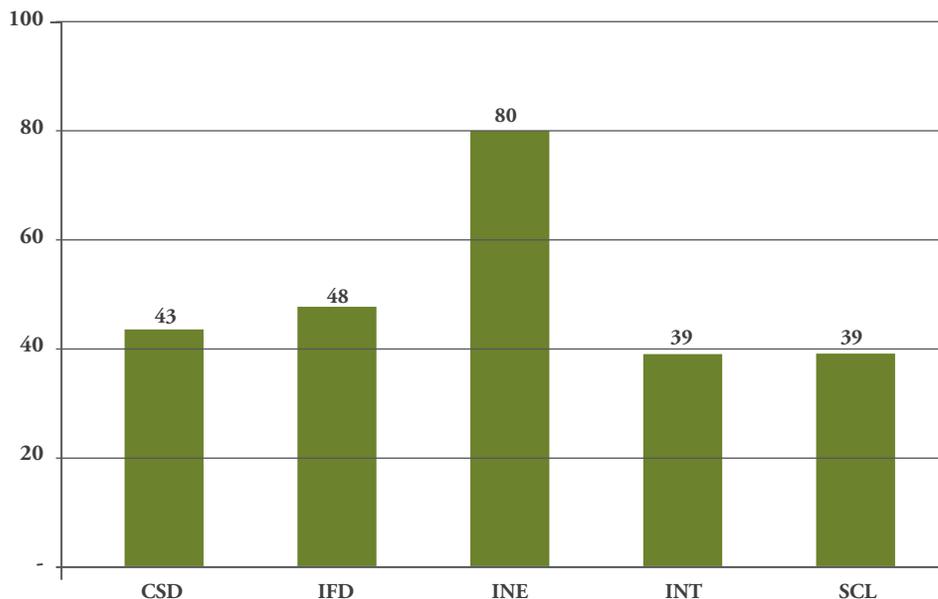
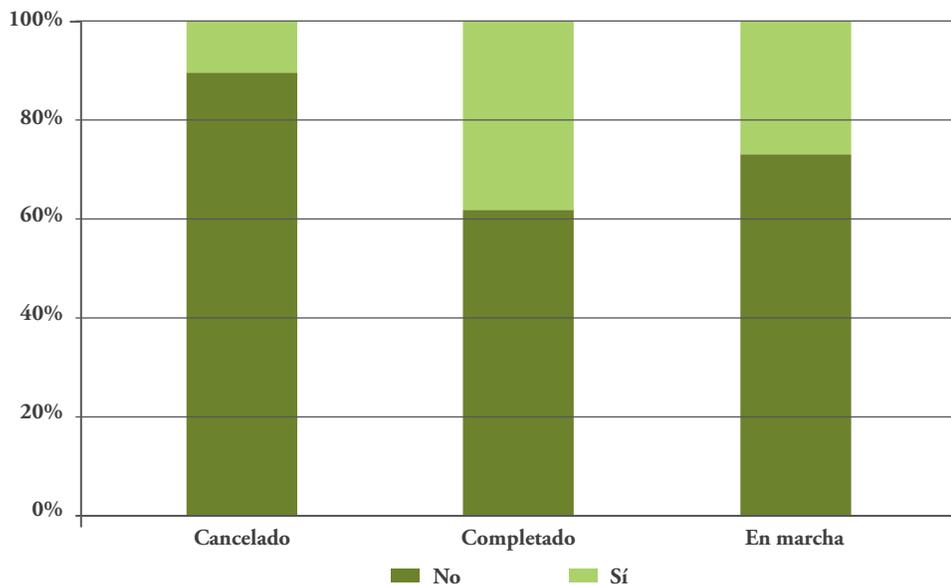


GRÁFICO 2.13
Uso de asociaciones para diseñar e implementar las evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



El Banco ha dependido enormemente de consultores y empresas consultoras para la producción de evaluaciones de impacto, si bien se han establecido pocas asociaciones para promoverlas (Gráfico 2.13)²⁹. Solo el 23% de las evaluaciones de impacto en marcha se están realizando en asociación con una institución externa, ya sea en lo que respecta al diseño o la implementación, lo cual representa una reducción sustancial frente al 37% de las evaluaciones de impacto completadas que se benefició de una colaboración externa³⁰. De las evaluaciones de impacto canceladas, solo el 10% se hizo en colaboración con otras instituciones. De todas las evaluaciones de impacto, solo el 1% se realizó o se está realizando en su totalidad a cargo de entidades asociadas, sin la participación directa del BID. Esto representa una oportunidad perdida para ampliar

el alcance de las evaluaciones de impacto, el uso de alternativas financieras o la difusión de los resultados. Otras instituciones han manifestado mayor interés en colaborar con asociados externos³¹.

En resumen, a pesar del avance alcanzado con el tiempo, el Banco todavía necesita fortalecer el proceso de originación de las evaluaciones de impacto. Los documentos de marco sectorial brindan una orientación heterogénea sobre las brechas de conocimiento, debido en parte al nivel de desarrollo del sector en cuanto a labor analítica. Muchas evaluaciones de impacto siguen siendo impulsadas por el Banco y no cuentan con el pleno compromiso del cliente. Los sectores tienen diferentes niveles de pericia y experiencia, aunque los que normalmente no han participado mucho en las evaluaciones de impacto se han vuelto considerablemente más activos. SPD ha ayudado en cierta medida al apoyar a sectores con menos experiencia, al menos en la etapa de diseño.

3. ¿Cuánto cuestan las evaluaciones de impacto?

En el BID las evaluaciones de impacto pueden financiarse a través de cuatro cauces.

- *Componentes de préstamo.* Cada proyecto tiene un presupuesto para seguimiento y evaluación, y algunos utilizan parte de esos recursos para financiar evaluaciones de impacto³².
- *Presupuesto administrativo del Banco.* Las divisiones del BID pueden presupuestar fondos para financiar evaluaciones de impacto correspondientes a operaciones completadas que no hayan recibido previamente una asignación de recursos o pueden complementar el financiamiento de un préstamo con presupuesto administrativo.
- *Fondos de operaciones de cooperación técnica.* Las operaciones de cooperación técnica pueden financiar evaluaciones de impacto de proyectos del BID (complementando el monto financiado por el préstamo o cubriendo todo el monto) o financiar evaluaciones que no estén directamente relacionadas con operaciones del Banco³³.
- *Recursos de actividades económicas y sectoriales.* Desde 2009, las evaluaciones de impacto pueden financiarse con recursos de actividades económicas y sectoriales.

La mayor parte de las evaluaciones de impacto consideradas en el presente análisis se financiaron mediante préstamos: 360 de las evaluaciones propuestas (68%) se incluyeron en un préstamo para inversión o en un préstamo en apoyo de reformas de política, en tanto que el resto se financió mediante operaciones de cooperación técnica o de financiamiento no reembolsable.

OVE observó una gran heterogeneidad en los presupuestos para evaluaciones de impacto, lo cual parece obedecer principalmente al tipo de datos empleados. OVE solicitó al personal encargado de las evaluaciones de impacto datos presupuestarios, incluidos todos los costos salvo el tiempo del personal del Banco. Como se indica en el Cuadro 2.2,

el 57% de las evaluaciones de impacto tiene presupuestos de menos de US\$250.000 y el 9% tiene presupuestado más de US\$1 millón. El principal factor que determina el costo no es la metodología en sí, sino el tipo de datos empleados en la evaluación (administrativos o recopilados). El costo promedio de una evaluación de impacto basada en datos administrativos únicamente asciende a US\$74.000, frente a US\$468.000 de las evaluaciones de impacto para las que se recopilan datos. Según las entrevistas, entre los principales factores determinantes del costo se incluyen el grado de dispersión de los beneficiarios y el país en que se lleva a cabo la evaluación. Es más probable que los países con sistemas de información deficientes recopilen datos para sus evaluaciones de impacto³⁴. Es importante destacar que los beneficios de recopilar datos pueden ir más allá de la evaluación de impacto en sí, especialmente en países con fuentes de información limitadas.

También se observa heterogeneidad entre los sectores en la asignación de recursos de cooperación técnica (Cuadro 2.2). IFD e INE han utilizado un monto limitado de recursos de los programas especiales del Capital Ordinario o de fondos fiduciarios para financiar evaluaciones de impacto, en tanto que SCL es el sector que emplea, por un gran margen, la mayor proporción de esos recursos. La distribución es desigual incluso dentro de SCL, observándose que LMK y GDI emplean menos recursos que los demás sectores.

CUADRO 2.2. DISTRIBUCIÓN DE LAS EVALUACIONES DE IMPACTO, POR PRESUPUESTO Y TIPO DE DATOS

Departamento	División	Total Evaluaciones de impacto. Método			Total Presupuesto (US\$ millones)				Costo promedio de la evaluación de impacto (millones de US\$)
		Ensayos aleatorios controlados	cuasi-experimental	No se define*	Coop. téc./ financ. no reembol., Capital Ordinario	Coop. téc./ financ. no reembol., fondos fiduciarios	Préstamo	Total	
CSD	CCS	1	2	-	-	1,0	1,0	1,0	0,65
	RND	1	17	2	1,7	4,5	18,8	20,5	0,47
	HUD	10	42	1	0,4	-	5,2	5,6	0,28
IFD	CMF	5	17	2	0,0	0,1	4,6	4,6	0,20
	CTI	5	33	9	-	0,3	9,4	9,4	0,21
	FMM	-	5	2	-	-	1,1	1,1	0,15
	ICS	21	39	12	1,9	0,8	7,7	9,6	0,14
INE	ENE	1	3	-	-	-	2,0	2,0	0,50
	TSP	3	11	-	0,2	-	9,9	10,1	0,72
	WSA	6	2	2	0,1	0,9	2,0	2,1	0,29
INT	TIN	1	21	1	0,1	-	0,6	0,7	0,03
SCL	EDU*	43	25	10	5,4	6,4	28,7	40,5	0,55
	GDI	11	5	4	1,1	3,1	1,0	2,1	0,26
	LMK	11	8	3	1,8	0,8	4,8	6,6	0,34
	SPH	49	40	27	5,3	7,3	54,6	59,9	0,58
	SCL	8	9	1	10,9	0,1	0,4	11,2	0,63
Total		176	279	76	28,8	25,2	151,8	187,0	387,6

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

Nota: *Se refiere a una evaluación de impacto incluida en los documentos de préstamo o los perfiles de las operaciones de cooperación técnica, pero cuya metodología no se definió completamente en el anexo de seguimiento y evaluación o en la matriz de productos.

OVE comparó el costo de las evaluaciones del BID con las cifras indicadas en la literatura especializada y las observadas en otros organismos y concluyó que el costo se sitúa dentro de una gama razonable. Gertler et al. (2016) intentaron establecer un costo de referencia de las evaluaciones de impacto analizando proyectos respaldados por el Fondo Estratégico para la Evaluación del Impacto (SIEF) que administra el Banco Mundial. Al igual que OVE, estos autores observaron que el costo varía ampliamente (entre US\$130.000 y US\$2,78 millones) y fluctúa de acuerdo con el tipo de evaluación y su magnitud. Al examinar el costo de las evaluaciones de impacto de otras organizaciones, OVE encontró que si bien la distribución presupuestaria del BID parece ser comparable a los montos que DIME y J-PAL gastan en sus evaluaciones, el costo es muy inferior al promedio de Millennium Challenge Corporation³⁵. El promedio de DIME es más alto porque depende considerablemente de experimentos y la recopilación de datos. Millennium Challenge Corporation también financia un mayor número de evaluaciones de impacto para las que se recopilan datos, y el costo de sus evaluaciones asciende a entre 3% y 5% del costo del proyecto.

En general, durante el período objeto de análisis, se presupuestó un total de US\$206 millones para financiar evaluaciones de impacto —US\$152 millones de operaciones de préstamo y financiamiento no reembolsable y los US\$54 millones restantes de operaciones de cooperación técnica (Cuadro 2.3)—. En el caso de las operaciones que incluyeron al menos una evaluación de impacto, el monto asignado a la evaluación promedia menos de 0,5% del monto total aprobado para la operación (Gráfico 2.14). Es importante señalar que esas estimaciones se basan en documentación disponible y la información brindada por los jefes de equipo (no en registros administrativos) y no incluyen los costos de las evaluaciones de impacto canceladas ni el tiempo de personal del Banco.

CUADRO 2.3. PRESUPUESTO TOTAL, POR INSTRUMENTO Y SITUACIÓN DE LA EVALUACIÓN DE IMPACTO (MILLONES DE US\$)

Evaluación de impacto	Coop.téc.	Préstamo	Total
Completada	19,2	28,9	48,1
Cancelada	10,9	45,4	56,3
En marcha	23,9	77,5	101,4
Total	54,0	151,8	205,8

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

El financiamiento no reembolsable concedido por el BID a evaluaciones de impacto asciende a 60% de lo que los países destinaron a esas evaluaciones por medio de préstamos³⁶. Los fondos no reembolsables dedicados a evaluaciones de impacto se cifran en US\$77 millones³⁷, lo que equivale a aproximadamente US\$0,61 por cada US\$1 invertido por los países (Gráfico 2.15). Sin embargo, la distribución por país es heterogénea y no se han establecido criterios claros de asignación. Paraguay, Chile, República Dominicana, Jamaica, México, Bahamas y Honduras recibieron un monto no reembolsable aproximadamente equivalente al nivel promedio. Otros países han recibido proporcionalmente más recursos no reembolsables por cada US\$1 de los préstamos: Venezuela (US\$5,25), Costa Rica (US\$4,98), Ecuador

(US\$1,71), Guatemala (US\$1,30) y Jamaica (US\$0,66). Los demás países recibieron menos de US\$0,61 por US\$1. Muchos especialistas indicaron durante las entrevistas que las solicitudes de financiamiento no reembolsable se debían al programa de investigación de una unidad y no coincidía necesariamente con las prioridades establecidas por el país. También es posible que las necesidades y los intereses de sectores del Banco y especialistas particulares influyan en las asignaciones.

GRÁFICO 2.14
Distribución de las evaluaciones de impacto, por fuente de financiamiento

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

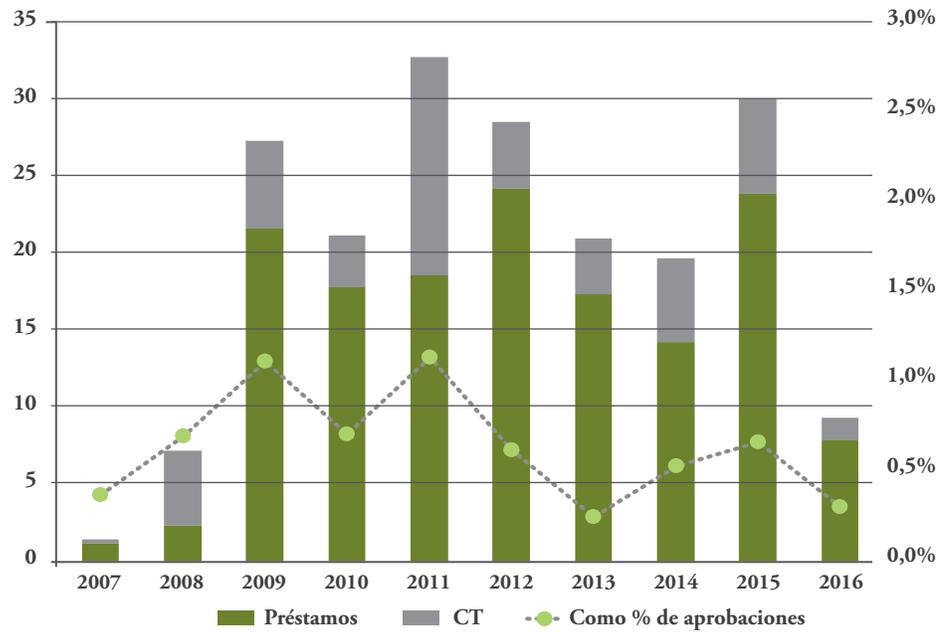
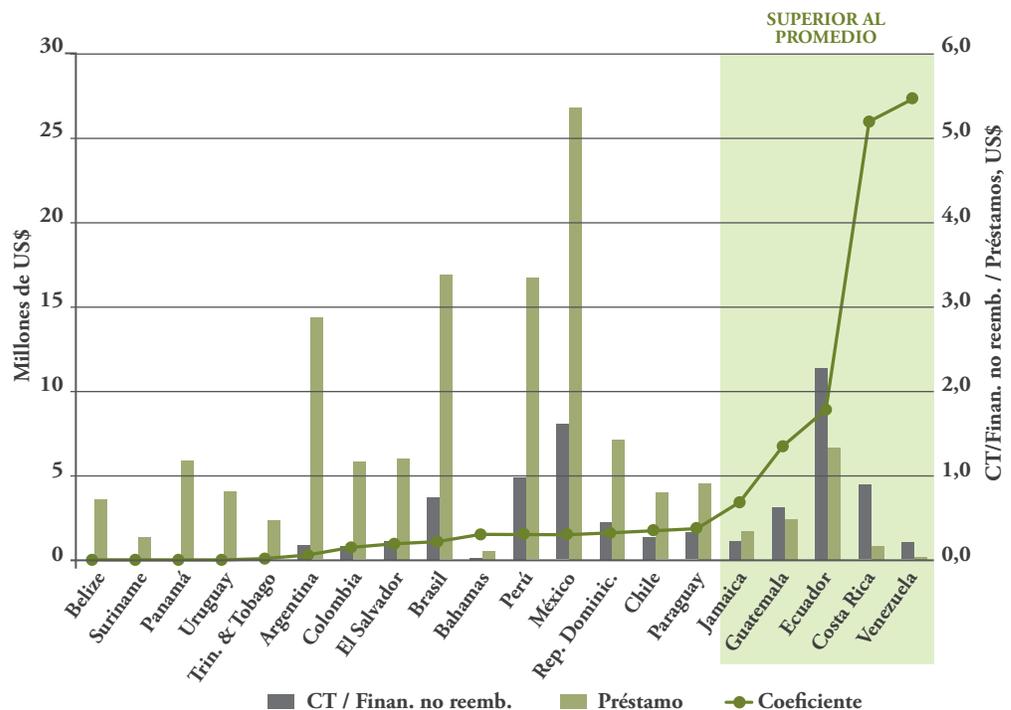


GRÁFICO 2.15
Evaluaciones de impacto, por presupuesto y país: operaciones de cooperación técnica, préstamos y coeficiente

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



En resumen, a lo largo de los años el Banco y los clientes han destinado una gran cantidad de recursos al financiamiento de las actividades de conocimiento, concretamente, las evaluaciones de impacto, y han acumulado una gran experiencia en la materia. Sin embargo, no parece haber habido mucho pensamiento estratégico a la hora de seleccionar las evaluaciones que se realizarían o asignar los recursos de cooperación técnica para financiarlas. En la siguiente sección se examinan algunas buenas prácticas en esta materia que se basan en la experiencia de varias organizaciones.

4. ¿Cuáles son las buenas prácticas en materia de financiamiento de las evaluaciones de impacto y los productos de conocimiento en general?

Los bancos multilaterales de desarrollo y otros interesados internacionales en el desarrollo están cada vez más comprometidos a producir y utilizar las evaluaciones de impacto y han recurrido a una variedad de mecanismos institucionales para financiarlas. Si se considera la experiencia propia del BID, así como la que han acumulado el Banco Mundial, la Organización para la Cooperación y el Desarrollo Económicos (OCDE), J-PAL y la Corporación Andina de Fomento (CAF), pareciera que tres factores básicos promueven los principios de eficacia y eficiencia en el financiamiento de las evaluaciones de impacto, a saber: la adopción de mecanismos claros de priorización, la participación de los clientes y la existencia de redes cooperativas.

Un proceso de priorización transparente y estratégico permite a los bancos multilaterales de desarrollo centrar la atención en brechas de conocimiento en sectores específicos, facilita el apalancamiento de recursos y reduce los costos de la puesta en práctica y divulgación de los conocimientos porque promueve las relaciones a largo plazo. Los donantes están más dispuestos a contribuir recursos cuando las reglas de asignación son transparentes. Asimismo, la transparencia en la priorización (como ocurre en DIME) ayuda a evitar la tendencia de los donantes de reservar recursos. Los bancos multilaterales de desarrollo no pueden producir todos los conocimientos internamente y muchas veces subcontratan productos específicos. Los modelos de producción de evaluaciones de impacto que son exitosos suelen suponer el establecimiento de vínculos sólidos con otros productores de conocimiento, como universidades y centros de investigación especializados, los cuales reportan ventajas reputacionales importantes para ambas partes y otros interesados dentro de una estructura de repartición de costos. Asimismo, el establecimiento de relaciones a largo plazo con otros proveedores de conocimiento puede reducir los costos de priorización al incorporar procedimientos de selección, como ocurre en J PAL y la Red de Centros del BID.

En los modelos exitosos de financiamiento de evaluaciones de impacto se observa una participación activa de los usuarios finales del conocimiento. Los proveedores de conocimiento deben tener un fuerte vínculo con los clientes y hacerlos participar realmente desde la fase de diseño de las evaluaciones de impacto. Ese vínculo tiene varios propósitos que contribuyen a las prácticas óptimas de financiamiento de evaluaciones de impacto: permite establecer estructuras para compartir costos, crea capacidad y

contribuye a los esfuerzos de difusión una vez concluida la evaluación. La repartición de costos es un elemento importante para armonizar los incentivos entre los diferentes interesados; crea un sentido de propiedad, al tiempo que permite que todas las partes obtengan los ahorros percibidos. En DIME, por ejemplo, los gobiernos participan desde el inicio del proceso y pagan el 30% del costo de las evaluaciones de impacto. Las contrapartes gubernamentales reciben capacitación sobre los métodos utilizados en las evaluaciones de impacto y participan activamente en su formulación e implementación. Ello promueve un sentido de propiedad entre las contrapartes gubernamentales, crea un mecanismo natural de control y equilibrio e incentiva la eficiencia en el uso de los fondos. Asimismo, es más fácil difundir una evaluación de impacto cuando el cliente está dispuesto a pagar por ella.

Por último, dado que el conocimiento es un bien público, establecer redes cooperativas facilita una mayor producción de evaluaciones de impacto. Los conocimientos se caracterizan por crecientes rendimientos de escala. Las redes cooperativas fomentan la difusión y el uso del conocimiento, que es necesario para incrementar los rendimientos, al tiempo que incentivan la especialización para desarrollar nuevos conocimientos. Asimismo, establecer sólidos mecanismos de cooperación tiene sentido desde el punto de vista de la eficiencia (se evitan duplicaciones y competencia). Si las evaluaciones de impacto incluyen un componente de subsidio, por ejemplo, la cooperación asegura que el precio de los diferentes proveedores se armonice, con lo cual se eliminan las oportunidades de arbitraje. Además de incrementar la eficiencia, las redes cooperativas permiten apalancar recursos porque aprovechan eficientemente los recursos en común y atraen aportes de terceros más eficazmente. A través de un ecosistema coherente de distintos tipos de redes (formales e informales) que cuente con temporalidad variada, diversos objetivos y personas de diferentes disciplinas, los bancos multilaterales de desarrollo podrían maximizar la creación de conocimiento e incrementar la tasa de rendimiento de sus inversiones.

Además de esas buenas prácticas generales, algunas organizaciones han establecido directrices y protocolos de investigación para definir cuándo y cómo deben realizarse evaluaciones de impacto. El Banco Mundial, por ejemplo, señala que una evaluación de impacto sería útil cuando (i) se está probando un mecanismo de intervención innovador, tal como un programa piloto; (ii) es probable que se amplíe la escala de la intervención o esta se reproduzca en un contexto diferente; (iii) la intervención es estratégicamente pertinente y requerirá un gran volumen de recursos; (iv) la intervención aún no se ha sometido a prueba o (v) los resultados de la intervención influirán en decisiones importantes de políticas³⁸. La mayor parte de los organismos que trabajan con evaluaciones de impacto cuentan con protocolos de investigación, y se ha demostrado que estos aumentan el interés de los donantes en patrocinar evaluaciones.

Si bien el BID ha estado al frente en muchos sentidos de la promoción de las evaluaciones de impacto, podría beneficiarse de la adopción de algunas de esas buenas prácticas. Tal como se señaló anteriormente, el Banco no cuenta con un sistema transparente

y estratégico para priorizar las evaluaciones de impacto. Para el Banco también podría ser útil promover una mayor participación del cliente y asegurar, en la medida de lo posible, su compromiso. Por último, si bien se han creado asociaciones para fines de capacitación (con UC Berkley y la Universidad de los Andes, entre otros) y en algunos casos para fines de implementación (por ejemplo con J-PAL), se podría lograr mucho más a través de un sistema más formal de asociaciones.

B. ¿CUÁLES EL NIVEL DE CALIDAD DE LAS EVALUACIONES COMPLETADAS Y DEL DISEÑO DE LAS QUE ESTÁN EN MARCHA?

Como primer paso en el análisis de la calidad de las evaluaciones de impacto del BID, OVE consideró el número de publicaciones generadas a partir de ellas. De las 94 evaluaciones de impacto completadas que se propusieron entre 2006 y 2016, más de la mitad no están documentadas en publicaciones. De las restantes, 15 se publicaron como parte de la serie de documentos de trabajo del BID, otras 15 se incluyeron en publicaciones especializadas y 9 se divulgaron por ambos medios. Ello significa que solo 39 han sido objeto de un proceso claro de control de calidad. Según las encuestas, muchos especialistas son renuentes a publicar en la serie de documentos de trabajo del Banco debido al rigor y el tiempo que supone ese proceso.

Consciente de que no existe una sola forma de evaluar la calidad de un producto analítico, OVE decidió indagar más sobre la calidad. La gran variación que existe de un sector a otro dificulta su comparación a lo largo del tiempo. El consenso que existe es que la mejor forma de evaluar la calidad es a través de un proceso de revisión de pares, en el que un grupo de profesionales evalúa diferentes propuestas en función de diferentes dimensiones, como la idoneidad de los datos, la metodología empleada y el aporte a la literatura especializada³⁹.

A fin de revisar más a fondo la calidad de las evaluaciones de impacto elaboradas por el BID, OVE contrató a CLEAR-LAC⁴⁰ y creó un marco de calidad que resume los principales elementos de las evaluaciones de impacto identificadas⁴¹. CLEAR-LAC sometió una muestra de 86 evaluaciones completadas⁴² y 59 propuestas de evaluaciones en marcha⁴³ a una revisión de pares, proporcionando en cada caso una evaluación global de la calidad y señalando las posibilidades de publicación en una revista especializada o como documento de trabajo (en el Anexo III figuran más detalles sobre la muestra seleccionada, el método empleado y los hallazgos de CLEAR). La evaluación de la calidad se basó en cuatro dimensiones.

- *Pertinencia:* La calidad de la revisión de la literatura especializada, la pertinencia de las preguntas planteadas en la evaluación y la calidad del protocolo de investigación.
- *Metodología:* La selección, el rigor y las deficiencias del método analítico, el momento en que se estableció el período postratamiento y de seguimiento, y el uso de cálculos del poder estadístico.

- *Datos:* Tamaño de la muestra, unidad de análisis y tipo y calidad de los datos.
- *Resultados:* El resultado del análisis y los problemas de implementación, como incumplimiento, abandono y efectos secundarios, así como las posibles repercusiones en la fiabilidad de la evaluación.

Según la evaluación de CLEAR, la calidad del 55% de las evaluaciones de impacto completadas es general o parcialmente satisfactoria, en tanto que la calidad del 45% restante es insatisfactoria o parcialmente insatisfactoria. Son numerosos los factores que inciden en esa evaluación global. Solo el 17% de las evaluaciones de impacto completadas incluye una revisión exhaustiva de la literatura especializada y un protocolo de investigación de buena calidad. Asimismo, solo el 42% de las evaluaciones de impacto completadas emplea enfoques metodológicos rigurosos, en tanto que el 23% presenta deficiencias metodológicas graves. En este sentido, es interesante destacar que no todos los métodos se aplican con el mismo rigor. Los ensayos aleatorios controlados son el enfoque metodológico más común y se emplean en el 35% de las evaluaciones de impacto. En las otras se emplean otros métodos cuasiexperimentales (56%) o métodos múltiples (9%). Aunque se observó que el 60% de los ensayos aleatorios controlados eran metodológicamente rigurosos, solo el 30% de las evaluaciones de impacto cuasiexperimentales mostró el mismo grado de rigor. Asimismo, los ensayos aleatorios controlados dan lugar a datos de mejor calidad que otros métodos: el 70% de los ensayos aleatorios controlados, pero solo el 55% de los otros métodos, incluyen descripciones detalladas sobre la recopilación de datos y la formulación de las variables. Si bien el 72% de las evaluaciones de impacto completadas tiene un tamaño muestral satisfactorio o parcialmente satisfactorio, el 78% no incluye o no explica los cálculos del poder estadístico; esa práctica es generalizada en todas las divisiones. Por último, la evaluación señala que las consideraciones relacionadas con la implementación, como el abandono o los efectos secundarios, incidieron en la fiabilidad de al menos el 34% de las evaluaciones de impacto.

La calidad varió significativamente de un sector a otro (Gráfico 2.16). Como se indicó en el análisis de cartera, la mayor parte de las evaluaciones de impacto emprendidas en divisiones que tradicionalmente no hacían ese tipo de evaluaciones aún no han sido completadas. Tres sectores (FMM, TSP y CMF) completaron tres evaluaciones de impacto o menos durante el período en cuestión. Las diferencias en sus evaluaciones globales de la calidad reflejan la complejidad de diseñar y realizar evaluaciones de impacto en esos sectores. Ello pone de manifiesto la importancia de extraer lecciones de las evaluaciones de impacto que fracasan para orientar la producción futura de ese tipo de evaluaciones, sobre todo en sectores con menos experiencia. En el otro extremo del espectro, tres divisiones (LMK, SPH y EDU) produjeron 14 evaluaciones de impacto o más durante ese período. Estas divisiones presentan una distribución más equilibrada en cuanto a la calidad, observándose que, en la división con el mejor desempeño, LMK, solo una evaluación de impacto se calificó como de calidad insatisfactoria. Las divisiones que producen un volumen mediano de evaluaciones de

impacto (GDI, RND, CTI e ICS) tuvieron, en promedio, el menor porcentaje de evaluaciones de impacto de calidad insatisfactoria y el mayor porcentaje de evaluaciones de calidad satisfactoria o parcialmente satisfactoria. Ello podría guardar relación con el uso de mejores mecanismos de priorización, dadas las limitaciones de capacidad para generar evaluaciones de impacto frente a los que producen un gran volumen de esas evaluaciones. OVE entrevistó a todos los jefes de división para comprender mejor los procesos de producción de las evaluaciones de impacto en cada división. En GDI y RND —divisiones con los mayores porcentajes de evaluaciones de impacto satisfactorias o parcialmente satisfactorias— los hallazgos de OVE corroboran la hipótesis de priorización.

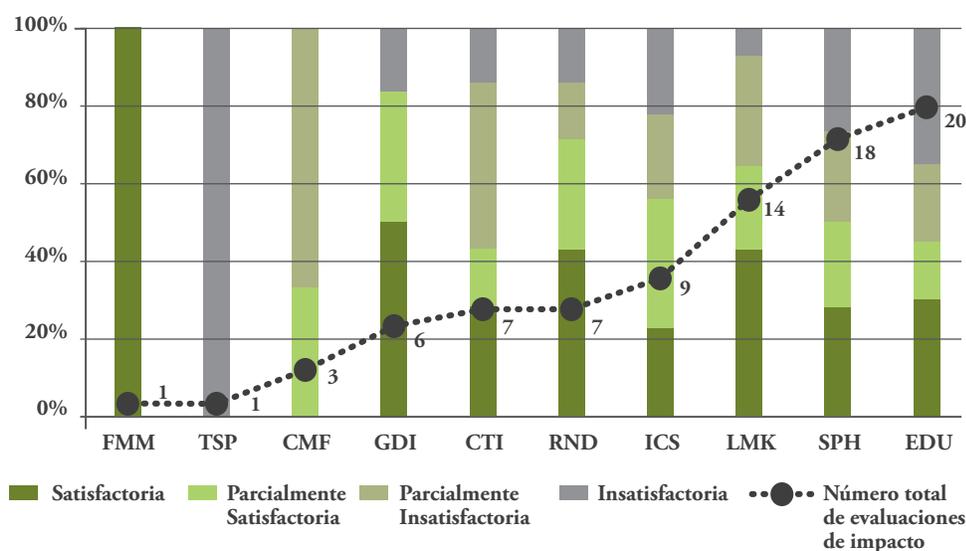


GRÁFICO 2.16
Calidad de las evaluaciones de impacto, por división

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.
Nota: En este gráfico no se incluyen 2014 y 2015 debido al pequeño número de evaluaciones de impacto propuestas en esos años que ya se completaron.

Si bien es difícil desentrañar las tendencias de calidad a lo largo del tiempo, puede afirmarse que la calidad de las evaluaciones de impacto ha ido mejorando. En el Gráfico 2.17 se indican los efectos de la DEM sobre la calidad. Antes de 2009, ciertos sectores, como SPH y EDU, estaban mucho más avanzados en el diseño e implementación de evaluaciones de impacto de alta calidad, y eran los únicos que incluían evaluaciones de impacto en sus operaciones de préstamo y de cooperación técnica. Sin embargo, la adopción de la DEM en 2009 creó grandes incentivos para producir evaluaciones de impacto e hizo que aparecieran “nuevos” sectores en el mercado. Ello llevó a una reducción de la calidad global de las evaluaciones de impacto propuestas entre 2009 y 2010, debido principalmente al menor grado de experiencia de algunos sectores. A partir de 2010, parece que la calidad tiende a mejorar, lo que demuestra que en general está aumentando la experiencia de los sectores en materia de evaluaciones de impacto. Asimismo, en una muestra representativa de propuestas⁴⁴ se observa que la calidad ha mejorado con el tiempo (Gráfico 2.18), pues la revisión de la literatura especializada es de mejor calidad, los protocolos de investigación son más rigurosos y los métodos analíticos son mejores. La mejora de la calidad de las propuestas es un paso positivo, ya que es el mejor indicador de la calidad general del producto final.

GRÁFICO 2.17
Evaluaciones de impacto completadas: evaluación de la calidad, por año

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

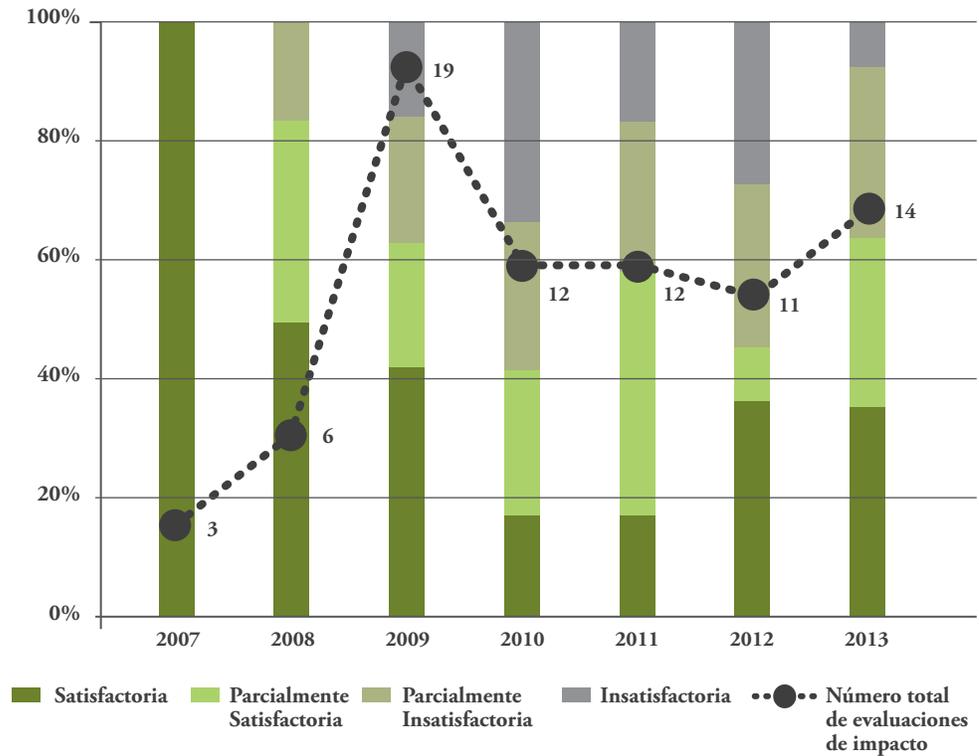
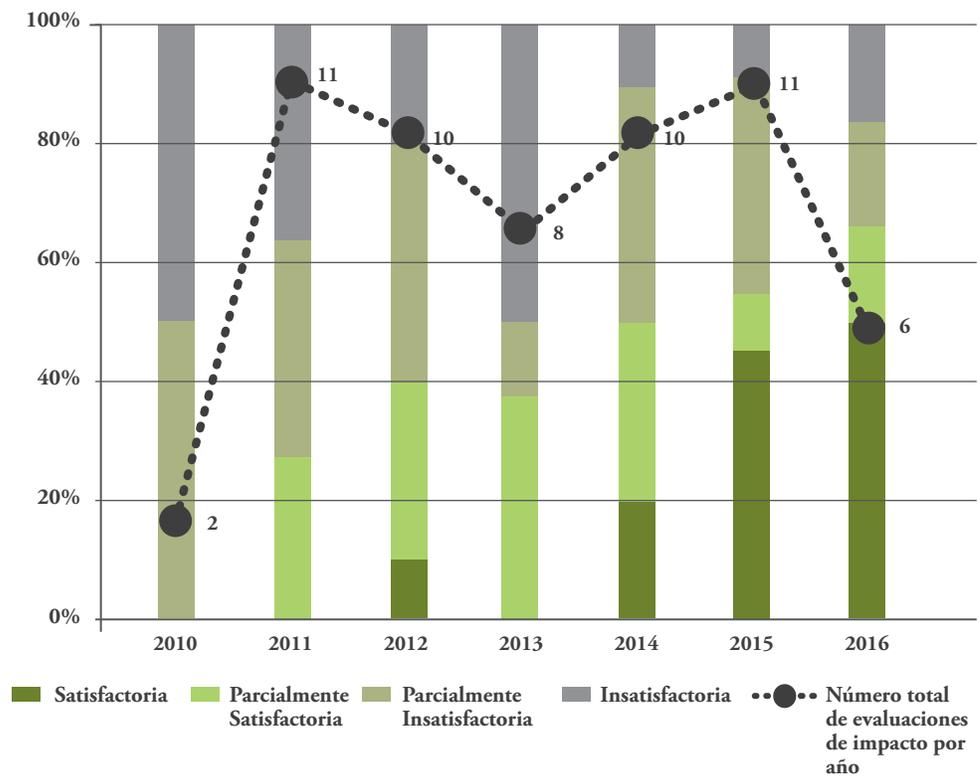


GRÁFICO 2.18
Propuestas de evaluaciones de impacto: evaluación de la calidad, por año

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



Este aumento de la calidad podría atribuirse a varios factores, incluidos en cierta medida el efecto del aprendizaje práctico y el gran esfuerzo realizado en capacitación sobre evaluaciones de impacto. SPD constituyó un centro de evaluación que cuenta con un impresionante volumen de material sobre evaluaciones de impacto (para todas las fases, desde el diseño hasta la difusión) y ha organizado amplias sesiones de capacitación para el personal del BID y sus clientes. Solo en 2015 y 2016, SPD/DSV dictó 40 cursos y talleres y organizó casi 50 más que contaron con la asistencia de más de 2.000 personas⁴⁵.

Uno de los principales motivos que impulsó la elaboración de evaluaciones de impacto es la mejora en la capacidad del Banco para promover la efectividad en el desarrollo en la región.

3 Uso e Influencia de las Evaluaciones del BID

El presente capítulo se centra en el uso y la influencia de las evaluaciones de impacto del BID. Es importante tener en cuenta que este tipo de análisis es generalmente muy complejo porque no hay una forma única de medir el uso y la influencia. Como forma de conocimiento, las evaluaciones de impacto pueden utilizarse para (i) orientar la labor y las estrategias del BID; (ii) orientar a los encargados de la formulación de políticas de la región en la toma de decisiones de políticas bien fundamentadas; y (iii) ayudar a cerrar brechas generales de conocimiento a fin de llegar a un público más amplio. En este capítulo se analizan cada uno de esos usos y una posible externalidad positiva de las evaluaciones de impacto, a saber, su uso como instrumento de apoyo al diseño y la implementación de proyectos.

A. USO E INFLUENCIA DE LAS EVALUACIONES DE IMPACTO COMO PRODUCTO DE CONOCIMIENTO

1. ¿En qué medida se usan las evaluaciones de impacto en el BID?

Uno de los principales motivos que impulsó la elaboración de evaluaciones de impacto es la mejora en la capacidad del Banco para promover la efectividad en el desarrollo en la región. En este sentido, es importante que las evaluaciones de impacto se utilicen internamente. OVE intentó comprender cómo emplea el personal del BID las evaluaciones y cómo las está integrando en el ciclo de proyectos y el proceso de programación. Esta evaluación plantea retos porque las evaluaciones de impacto sirven propósitos diferentes y, simultáneamente, se dirigen a públicos diferentes, y hasta la fecha no se han realizado tantas. OVE realizó entrevistas semiestructuradas a funcionarios de VPC y VPS para determinar el uso que le da el personal que interviene

en las operaciones del Banco a los resultados de las evaluaciones. Las entrevistas dejaron ver varias percepciones sobre la forma en que las evaluaciones de impacto se usan en la práctica y su importancia relativa en comparación con otros productos analíticos.

Según un análisis de los documentos de marco sectorial, el BID no emplea las evaluaciones de impacto sistemáticamente para orientar o ayudar a definir enfoques generales a nivel sectorial. El uso de las pruebas empíricas que generan las evaluaciones de impacto puede mejorar la teoría del cambio en que se sustentan las operaciones del BID, promover el aprendizaje institucional, enriquecer la estrategia para lograr la participación de los interesados y servir de base para el diseño de la labor de seguimiento continuo. Si bien los documentos de marco sectorial no deben específicamente tener en cuenta y orientar las evaluaciones de impacto, sí deben incluir toda información que sea pertinente para mejorar la calidad de la labor operativa y analítica del Banco a nivel sectorial. Asimismo, esos documentos se actualizan periódicamente (cada tres años) para promover el aprendizaje continuo, lo que los convierte en el medio natural por el cual se transmiten las pruebas empíricas que van surgiendo de las evaluaciones de impacto a la toma de decisiones. No obstante, el 40% de los documentos de marco sectorial vigentes no incluyen pertinentemente las evaluaciones de impacto cuando citan pruebas empíricas internacionales a nivel sectorial⁴⁶.

Las entrevistas revelaron que hay dos etapas en el ciclo de préstamos en que las evaluaciones de impacto tienen mayores probabilidades de ser utilizadas: La primera, durante el proceso de programación, cuando se discute la programación de préstamos con el país y, luego, durante la etapa de diseño del proyecto, cuando se planifican intervenciones específicas en detalle. En ambos puntos son de utilidad la experiencia de operaciones anteriores y los hallazgos de la literatura especializada basada en pruebas empíricas.

Durante el proceso de programación de país, los resultados de programas implementados parecen emplearse junto con otros documentos basados en pruebas empíricas. Las estrategias de país brindan la base de la programación futura. El proceso ideal de elaboración de las estrategias tarda unos dos años: comienza un año antes del cambio de gobierno en el país en cuestión y concluye un año después. En general, las discusiones durante el primer año se centran en el papel que podría ejercer el BID en los diversos sectores del país y, en el segundo año, se establece la programación final de préstamos. Varios instrumentos analíticos (informe diagnóstico del crecimiento, evaluación del programa de país y notas de políticas y sectoriales/temáticas) ayudan a diagnosticar las necesidades del país y brindan los antecedentes técnicos para las negociaciones sobre temas específicos. OVE examinó las fuentes de las pruebas empíricas utilizadas en las estrategias de país y las notas sectoriales/temáticas y halló que las evaluaciones de impacto no se citan frecuentemente. Algunos sectores (SCL, INT) emplean evaluaciones de impacto con mayor frecuencia, junto con literatura especializada basada en pruebas empíricas de publicaciones académicas, en tanto que otros (INE, ICF) emplean principalmente otros insumos analíticos generados en cada proyecto.

Aunque las pruebas empíricas sobre el uso de las evaluaciones de impacto en las estrategias de país, las notas técnicas y los documentos de marco sectorial son limitadas, las entrevistas apuntan a que ocupan un lugar importante en el diálogo de país: para evitar programas de eficacia limitada. El apoyo del BID responde más que todo a la demanda, por lo que los intereses de cada país, que muchas veces son de tipo político, son sumamente pertinentes. En algunos casos, la presentación de pruebas empíricas sobre la eficacia puede ayudar a orientar a los gobiernos para que emprendan intervenciones diferentes y mejores. El programa “Una Laptop por Niño” es un caso patente. El programa —por un valor equivalente a alrededor de la cuarta parte del gasto general del BID en evaluaciones de impacto— se había puesto en marcha en algunos países cuando los resultados de la evaluación de impacto indicaron que no incidía en el rendimiento de los niños. En consecuencia, el programa se canceló o no fue aprobado en al menos tres países (Perú, Chile y México).

El uso de las pruebas empíricas de las evaluaciones de impacto del BID durante la fase de diseño de los proyectos sigue siendo muy heterogéneo. SCL e INT usan activamente los resultados de las evaluaciones de impacto, más que otras unidades sectoriales, por distintas razones. En primer lugar, solo un número relativamente pequeño de evaluaciones completadas abordan temas que son pertinentes para otros sectores. Segundo, muchas de las evaluaciones que sí existen son de calidad regular y no necesariamente se centran en las preguntas más importantes para el diseño de los proyectos. Tercero, se observa una cierta inercia en la forma de trabajar del Banco, pues muchas veces se utilizan diseños similares de proyectos en diferentes países, independientemente de la falta de pruebas sobre su eficacia o, peor todavía, a veces con pruebas de su ineficacia.

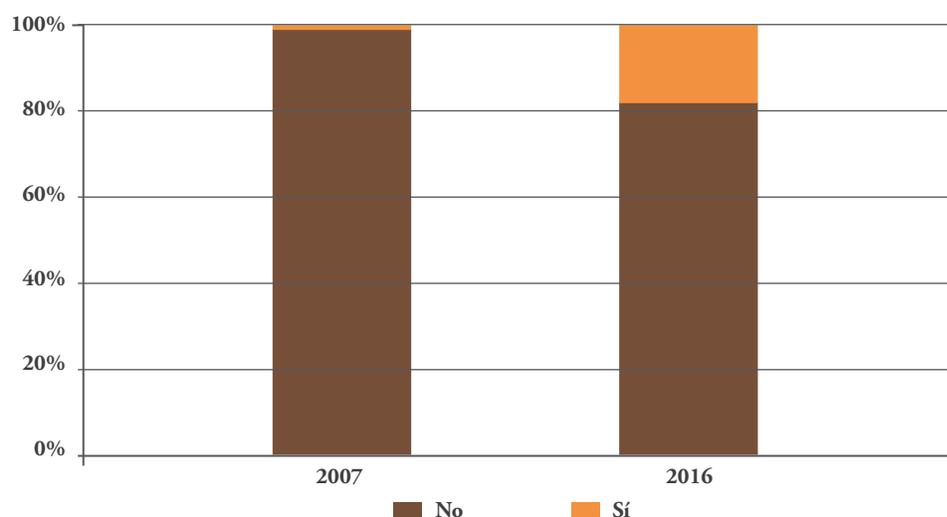


GRÁFICO 3.1
Operaciones cuya documentación cita resultados de evaluaciones de impacto para respaldar su diseño

Fuente: OVE. Análisis de publicaciones.

Además de las entrevistas, OVE analizó las referencias a evaluaciones de impacto en documentos de préstamo elaborados por el Banco y confirmó la forma heterogénea en que los sectores las emplean. OVE revisó todos los documentos de préstamo

aprobados en 2007 (antes de adoptarse el Marco de Efectividad en el Desarrollo) y en 2016 (el año más reciente para el cual se dispone de datos) en busca de citas a evaluaciones de impacto en la descripción del diseño de la intervención. Aunque los proyectos de SCL son los que tienen la mayor probabilidad de basarse en pruebas empíricas de evaluaciones de impacto, el análisis señala en general que todos los sectores aumentaron el uso de las evaluaciones de impacto en el diseño de proyectos (Gráfico 3.1)⁴⁷.

2. ¿Cómo emplean los clientes las evaluaciones de impacto del BID?

Las evaluaciones de impacto del Banco podrían desempeñar un papel fundamental en el proceso de formulación de políticas de los clientes, pero es difícil evaluarlo. Además de facilitar un proceso de retroalimentación en tiempo real para los encargados de formular políticas, podrían ayudar a definir el diseño más eficaz de programas futuros⁴⁸. Las evaluaciones de impacto también podrían utilizarse para conferir validez externa a los proyectos. Pueden ofrecer un punto de vista objetivo y proporcionan pruebas empíricas rigurosas que pueden ayudar a los responsables de la formulación de políticas a decidir, por ejemplo, si ampliar, modificar o eliminar un programa. A juicio de OVE, el mejor enfoque para evaluar el papel de las evaluaciones de impacto en la formulación de políticas sería medir el contexto de las modificaciones de las políticas públicas durante un período largo. Sin embargo, un enfoque de tal índole tiene dos limitaciones, a saber: (i) no puede aislarse la influencia de las evaluaciones de impacto del BID de otras fuentes de conocimiento y (ii) la producción de evaluaciones de impacto en el BID es todavía demasiado reciente. En la práctica, el principal proceso a más largo plazo que se observa es el programa PROGRESA/Oportunidades (Recuadro 3.1).

OVE entrevistó a una muestra de clientes del BID que participaron en proyectos con evaluaciones de impacto, y la mayoría de ellos señaló que su principal beneficio es la posibilidad de ajustar el funcionamiento de una política en marcha. OVE seleccionó una muestra estratificada, por departamento, de 50 evaluaciones de impacto (completadas y en marcha) para realizar el seguimiento con las contrapartes gubernamentales del Banco. Sin embargo, OVE logró contactar a al menos una persona del gobierno para la entrevista en solo la mitad de la muestra. El resultado fue una muestra de 25 evaluaciones de impacto⁴⁹. Un total de 52% de los entrevistados opinó que las evaluaciones de impacto son principalmente una herramienta para ajustar políticas, en tanto que 20% señaló que son un instrumento para apoyar y validar políticas vigentes. Curiosamente, algunos subrayaron la importancia de las evaluaciones de impacto como herramienta de rendición de cuentas para justificar la obtención de financiamiento del BID.

OVE constató una amplia gama de opiniones acerca de la utilidad de las evaluaciones de impacto. Alrededor de 56% de los entrevistados señaló que la evaluación de impacto ejerció influencia, o preveían que ejercería influencia, en la formulación de

políticas, en tanto que 40% afirmó que emplearon la evaluación, o preveían emplearla, en discusiones sobre otras políticas. Entre los entrevistados cuyas evaluaciones de impacto estaban en marcha (en lugar de completadas), el uso o la expectativa de uso en la formulación de políticas vigentes aumentó a 83%. Solo un entrevistado indicó que las evaluaciones de impacto no tenían valor y que representaban un costo; otro opinó que en general las evaluaciones de impacto no se llevan a cabo con la calidad necesaria para aislar los efectos y orientar las políticas. Los entrevistados que señalaron que las evaluaciones tenían poco o ningún valor para ajustar políticas vigentes citaron las consideraciones políticas y los cambios de gobierno como factores recurrentes. Asimismo, se identificaron como barreras a su uso temas relacionados con el momento en que se realizan las evaluaciones de impacto (en el Recuadro 3.2 se presenta un análisis más amplio de los factores que limitan el uso).

Recuadro 3.1. Una sola evaluación de impacto no produce el cambio: El caso de PROGRESA/Oportunidades

PROGRESA/Oportunidades se considera como un ejemplo de una política basada en resultados. El programa fue objeto de una evaluación rigurosa y el modelo se aplicó en otros países. La evaluación del programa sirvió para incrementar el costo de discontinuar el programa cuando un gobierno de la oposición asumió el poder. Asimismo, sirvió para conferir legitimidad a la intervención. Ambas funciones van más allá del conocimiento.

No obstante, las pruebas empíricas de fuentes distintas a las evaluaciones de impacto tuvieron una función igualmente pertinente en la formulación del programa. Las decisiones sobre la estructuración de los pagos, las condiciones y los desembolsos a las madres —entre muchas otras— se sustentaron en pruebas empíricas, recopiladas a lo largo de decenios, sobre la dinámica de la asignación de los recursos en los hogares. Ello significa que si bien las pruebas empíricas ejercieron un papel en PROGRESA, el mecanismo se basó en una serie de trabajos sobre desarrollo y no necesariamente en evaluaciones de impacto específicas. Lo que no está claro es la medida en que las evaluaciones de impacto dieron lugar a cambios en el programa. Behrman (2010) señala que las opiniones varían, pero que hay considerable escepticismo con respecto a qué cambios se vieron verdaderamente influenciados por la evaluación.

En cualquier caso, las evaluaciones de impacto de PROGRESA han tenido una función mucho más pertinente como bien público que en términos de la operación del programa. Según una revisión reciente, que incluyó amplias entrevistas a diseñadores y evaluadores de programas, la opinión unánime es que el programa de evaluaciones de impacto fue fundamental para reproducir el modelo en otros países y promover su continuidad en México. La estrategia de evaluación ayudó a hacer los programas de transferencias condicionadas de efectivo más aceptables y atractivos, y en muchos de los programas posteriores de transferencias condicionadas se ve la gran influencia de los programas anteriores, especialmente PROGRESA/Oportunidades. Cabe señalar que esa influencia se debió no tanto al contenido de las evaluaciones, sino al hecho de que se realizaron independientemente y a que se consideraron creíbles en su mayor parte.

Recuadro 3.2. Obstáculos al uso de las evaluaciones de impacto en la formulación de políticas

Pese a la posible utilidad de las evaluaciones de impacto en la formulación de políticas, son varios los motivos que explican por qué se subutilizan.

- Competencia con otras fuentes de pruebas empíricas (Ravallion, 2008). Las evaluaciones de impacto suelen redactarse con un alto grado de rigor académico y centrarse más en los métodos que en las repercusiones de políticas y, en consecuencia, resultan bastante difíciles de utilizar para un público no académico. En cierta medida, este problema se resuelve mediante la difusión estratégica de evaluaciones de impacto completadas. Al respecto, INT constituye un buen ejemplo: una vez que el sector completa varias evaluaciones, publica libros con el objetivo único de difundir las lecciones aprendidas (sin discusiones técnicas).
- Enfoque distinto al de formulación de políticas (Ravallion, 2008). Las evaluaciones de impacto suelen centrarse en la pregunta sobre qué funciona, en tanto que los encargados de formular políticas están más interesados en la eficacia en función de los costos. Para abordar ese asunto, debe hacerse un esfuerzo para que se realicen mejores evaluaciones que no se limiten a medir solo el impacto y que analicen el tema de la eficacia. El presente informe demuestra que el BID está avanzando en este sentido y que responde más las preguntas que interesan a los encargados de formular políticas, en lugar de simplemente emplear las evaluaciones de impacto como herramienta de rendición de cuentas para sus proyectos.
- Consideraciones relacionadas con el momento en que se realiza una evaluación de impacto (Sutcliffe y Court, 2005). Con frecuencia, los encargados de formular políticas deben asesorar a ministros sobre temas apremiantes. Lamentablemente, tanto la producción como la posterior publicación de los conocimientos científicos tarda tiempo y, debido a ese desfase, muchos encargados de formular políticas optan por emplear métodos o supuestos menos rigurosos o basarse en valores personales o institucionales, sin esperar pruebas empíricas sólidas. Debe tenerse en cuenta esta disyuntiva entre hallazgos normativos oportunos y pruebas empíricas rigurosas (sobre todo en los bancos de desarrollo); cuando finalmente aparecen las pruebas empíricas es posible que ya no sean pertinentes.

Los entrevistados tendieron a considerar las evaluaciones de impacto como útiles en los casos en que el gobierno había participado directamente en el diseño. Como sostienen Weyrauch y Díaz Langor (2013), la comunicación entre los agentes pertinentes es en efecto fundamental para determinar cómo la investigación llega a aplicarse en las políticas. El Banco parece estar internalizando este hecho e intensificando su comunicación con los clientes a fin de ajustar más las evaluaciones de impacto a las necesidades del cliente en materia de conocimiento. Por ejemplo, todos los entrevistados con evaluaciones de impacto en marcha afirmaron que estas se ajustaban a sus necesidades en cuanto a pruebas empíricas.

La creación de capacidad y el mayor interés en integrar las evaluaciones de impacto a los sistemas de seguimiento y evaluación de los gobiernos son dos externalidades positivas que citaron los entrevistados. Más de la mitad declaró que los requisitos de datos de las evaluaciones de impacto se tradujeron en mejoras en sus sistemas de evaluación y seguimiento o de presentación de información sobre las actividades. Asimismo, dos tercios de los entrevistados afirmaron que tras haber participado en la evaluación de impacto del BID estaban mejor preparados para realizar ese tipo de evaluaciones en el futuro, en parte por la capacitación recibida del Banco⁵⁰. Incluso los entrevistados que no recibieron capacitación señalaron que estaban mejor preparados para contratar a consultores para hacer el trabajo.

3. ¿Cuál es el alcance más amplio de las evaluaciones de impacto del BID?

Las evaluaciones de impacto solo pueden ejercer influencia en la medida en que se utilizan; pero para poder emplear los resultados, estos deben conocerse. De las evaluaciones de impacto del Banco completadas, 55 (58%) no se divulgaron ni en publicaciones especializadas ni en la serie de documentos de trabajo del BID⁵¹. Muchas ni siquiera pudieron consultarse en IDBDOCS y OVE tuvo que pedírselas directamente al jefe de equipo. Ello demuestra que los resultados de las evaluaciones de impacto del BID no son de fácil acceso para el público.

OVE también hizo un análisis de las citas en los artículos y libros que publica el Banco y comparó su frecuencia y patrones con los de otros bancos multilaterales de desarrollo⁵². Entre 2006 y 2016, autores afiliados al BID publicaron 177 artículos que citaban evaluaciones de impacto⁵³; estas evaluaciones de impacto se citaron un total de 2.603 veces, es decir un promedio de 16 citas por publicación y 237 citas por año de publicación. Los años en que se publicó el mayor número de evaluaciones de impacto fueron 2012 (23) y 2015 (27), y los años de publicación con el mayor número de citas fueron 2012 (515), 2013 (377) y 2011 (312). No obstante, como indica el Gráfico 3.2, el promedio de citas por publicación en el caso de las evaluaciones de impacto publicadas en 2015 (3) es muy inferior al promedio de evaluaciones publicadas en 2012 (26), quizás porque esas evaluaciones son más recientes y aún no aparecen en trabajos de investigación de seguimiento. Las citas se incluyen principalmente en notas técnicas del BID y la mayor parte son atribuibles a un pequeño número de autores. Durante ese período, los 14 autores más publicados representaron casi el 90% de las citas.

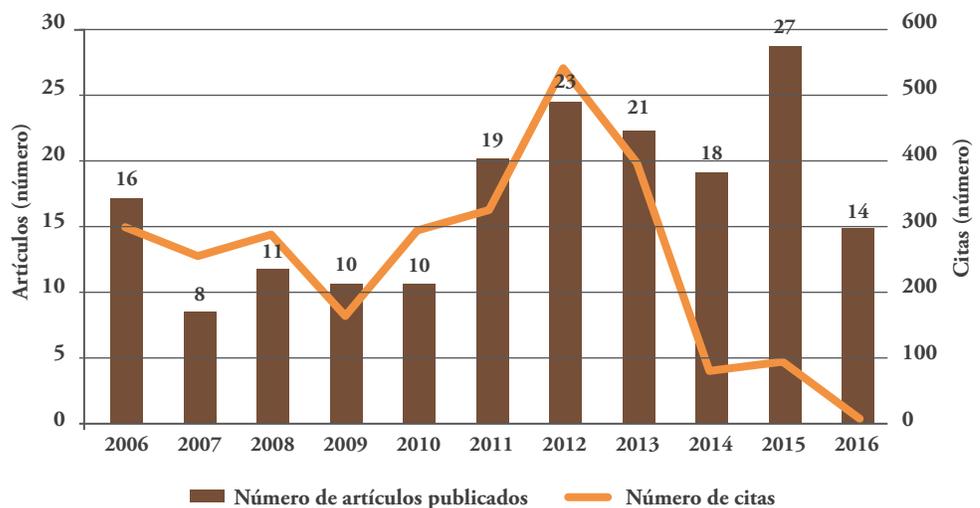
A fin de poner ese análisis en contexto, OVE examinó el número de citas de publicaciones generales del BID en relación con otros organismos y halló que, con respecto a sus comparadores, el BID tiene un desempeño promedio. En el caso del BID, el número promedio de citas por publicación es de 7,9 y corresponde a todas las publicaciones realizadas entre 1996 y 2010. Para ese mismo período de análisis, los promedios más bajos corresponden al Banco Asiático de Desarrollo (4,2), el Banco Africano de Desarrollo (3,4), el Banco Europeo de Inversiones



(6,8) y el Fondo Internacional de Desarrollo Agrícola (3,9), en tanto que los más altos son los del Banco Mundial (13,5), el Fondo Monetario Internacional (8,9), el Programa de las Naciones Unidas para el Desarrollo (18,6) y el Banco Europeo de Reconstrucción y Desarrollo (11,2)⁵⁴.

GRÁFICO 3.2
Publicación y citas de evaluaciones de impacto del BID, por año de publicación

Fuente: OVE, sobre la base de los resultados sobre citas del programa informático Publish or Perish.



B. ¿PUEDEN LAS EVALUACIONES DE IMPACTO FACILITAR LA EJECUCIÓN DE PROYECTOS?

La mayor parte de los esfuerzos dirigidos a incrementar la producción de evaluaciones de impacto en el BID obedeció a la idea de fortalecer el conocimiento y asegurar la rendición de cuentas. Sin embargo, hay pruebas empíricas que muestran que cuando se justifican y se diseñan adecuadamente las evaluaciones de impacto también pueden facilitar la ejecución de proyectos. Legovini, DiMaro y Piza (2015) se sustentan en datos sobre la ejecución de proyectos en el Banco Mundial para demostrar que los proyectos que tienen evaluaciones de impacto son objeto de desembolsos más rápidos. Los autores enumeran varios cauces a través de los cuales las evaluaciones de impacto podrían facilitar la ejecución de proyectos, por ejemplo, mediante mejoras en la planificación y el diseño basado en pruebas empíricas y una mayor capacidad de implementación gracias a la labor del equipo de investigación y el personal en el terreno. Asimismo, sostienen que el proceso de preparación de evaluaciones de impacto incrementa el volumen de datos para la toma de decisiones de políticas, lo cual ayuda al proyecto y genera un efecto observador y motivación, pues los conocimientos sobre una evaluación de impacto pueden generar expectativas.

OVE halló pruebas empíricas de que los proyectos del BID aprobados entre 2009 y 2016 que incluyeron una evaluación de impacto desembolsaron sus recursos más rápidamente que los proyectos que no la incluyeron⁵⁵. Se observa una clara diferencia entre las curvas de desembolso de los proyectos con y sin evaluación de impacto (Gráfico 3.3). Si bien parecen ser casi iguales durante los dos primeros años de ejecución, los proyectos con evaluación de impacto luego aceleran sus desembolsos y se completan antes. Aunque la duración prevista de proyectos con y sin evaluación de impacto es similar en promedio, los que no la incluyen tienen mayores retrasos (tres meses más en promedio). En el Anexo VII se presenta el análisis econométrico y se demuestra que la diferencia es significativa y robusta incluso cuando se neutralizan los efectos de muchas características de los proyectos, como el monto del préstamo, el sector, el país, el número de cambios de jefe de equipo, el año de aprobación e incluso el puntaje sobre la lógica del proyecto en la DEM (como variable sustitutiva de la calidad del proyecto).

No está claro cuáles son los cauces exactos a través de los cuales las evaluaciones de impacto podrían facilitar la ejecución de proyectos. OVE intentó arrojar luz sobre el tema comparando las tasas de desembolso de proyectos con evaluaciones de impacto completadas o en marcha y con posibilidad de completarse con proyectos que incluyen evaluaciones de impacto que se han cancelado o suspendido. El Gráfico 3.4 es análogo al Gráfico 3.3, con la diferencia de que se separan los proyectos con evaluaciones de impacto en dos categorías: (i) proyectos con evaluaciones de impacto

completadas o en marcha (los casos “exitosos”) y (ii) proyectos con evaluaciones de impacto canceladas o suspendidas (los casos “no exitosos”). Se observa una leve diferencia entre las curvas de desembolso de proyectos sin evaluaciones de impacto y las de proyectos con evaluaciones canceladas, pero ello también es válido para proyectos con evaluaciones de impacto completadas o en marcha. Esto respalda la idea de que las evaluaciones de impacto ejercen una función en la ejecución de proyectos, probablemente debido al respaldo y la capacitación brindados por los investigadores y el personal en el terreno asignados al proyecto. El respaldo gubernamental que requieren las evaluaciones de impacto también podría ser uno de los factores que explican esos resultados.

GRÁFICO 3.3
Desembolsos de los préstamos para inversión, con y sin evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.

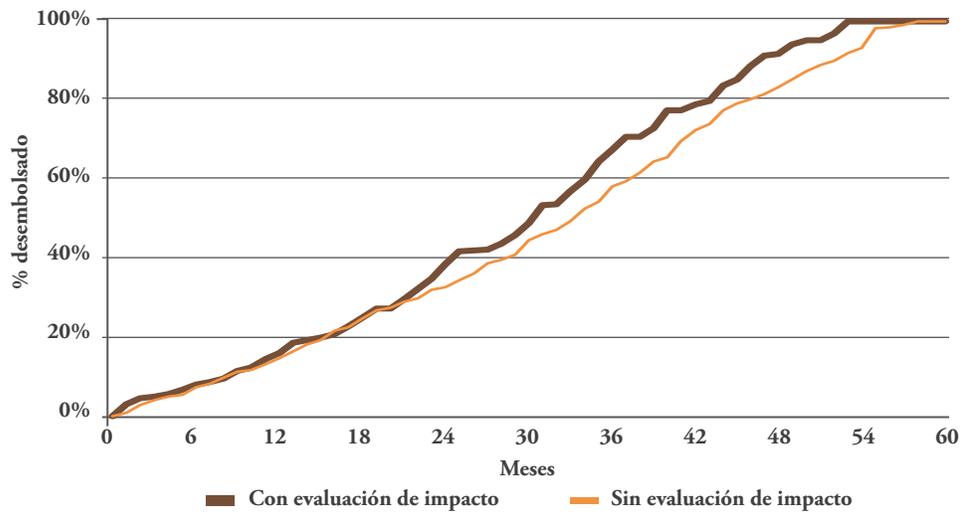
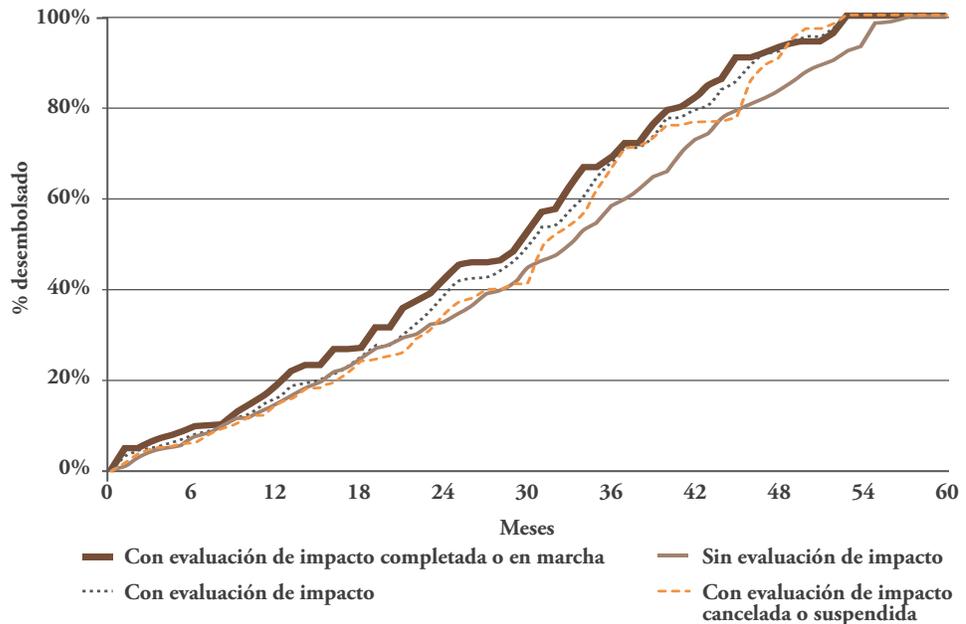


GRÁFICO 3.4
Desembolsos de préstamos para inversión, con y sin evaluaciones de impacto

Fuente: OVE. Base de datos de evaluaciones de impacto, 2017.



El posible rendimiento superior de un préstamo con evaluación de impacto tiene repercusiones financieras y podría contrarrestar hasta la mitad del costo de la evaluación⁵⁶. OVE estimó ese ahorro utilizando el monto promedio de los préstamos con y sin evaluaciones de impacto y las comisiones de crédito que se aplican actualmente a los saldos no desembolsados. El ahorro promedio producto de un desembolso más rápido es de unos US\$40.000 por año. Suponiendo un ciclo de proyecto de cinco años en promedio, el ahorro total para los clientes se sitúa entre 25% y 50% del costo promedio de la evaluación de impacto⁵⁷. Sustentándose en este análisis, OVE concluye que las evaluaciones de impacto cuando están bien justificadas pueden, además de complementar los conocimientos en algunos ámbitos, traer consigo externalidades positivas en la ejecución de los proyectos y la reducción de los costos financieros.



El Banco ha avanzado hacia la elaboración de evaluaciones de mejor calidad, pero todavía sigue habiendo margen para mejoras.

© BID

4 Conclusiones y Recomendaciones

El presente documento hace un balance del esfuerzo realizado para promover la práctica de producir evaluaciones de impacto en el BID. El primer hallazgo es la falta de información sistematizada sobre esas evaluaciones. Tras más de 10 años de inversión en evaluaciones de impacto, el Banco aún no cuenta con un sistema centralizado para llevar a cabo el seguimiento de todas las evaluaciones de impacto en marcha y su costo, ni dispone de un registro básico que incluya información sobre las preguntas de evaluación, el cumplimiento de normas éticas y los protocolos de investigación. SPD creó un sistema para hacer un seguimiento de las evaluaciones de impacto vinculadas a préstamos, y actualmente elabora un sistema para las que se incluyen en operaciones de cooperación técnica. No obstante, no es fácil dar seguimiento a las evaluaciones de impacto financiadas con recursos de actividades económicas y sectoriales, ni tampoco al costo exacto de las evaluaciones. Asimismo, aún no existe un repositorio central para archivar los informes o los conjuntos de datos.

Los incentivos creados por el Banco en los últimos años han hecho que el personal proponga un número excesivo de evaluaciones de impacto. El sistema de calificación de la DEM hizo que muchos equipos propusieran evaluaciones que no siempre cumplían las condiciones necesarias para la implementación. Cabe destacar dos consecuencias. En primer lugar, muchas evaluaciones fueron canceladas, lo que limitó la información sobre los resultados de proyectos con

mecanismos de seguimiento y evaluación deficientes. OVE no logró determinar el costo monetario de iniciar esas evaluaciones fracasadas. Segundo, los grandes incentivos creados por la DEM parecen haber fomentado un cambio cultural en el Banco. En general, las divisiones están haciendo un mayor hincapié en la efectividad en el desarrollo y en cómo evaluarla, si no es con evaluaciones de impacto rigurosas, al menos con una mentalidad de hipótesis alternativas. Asimismo, en las divisiones que llevan trabajando más tiempo con evaluaciones de impacto, se observa un claro proceso de aprendizaje. Actualmente esas divisiones parecen haber adoptado un enfoque más estratégico con respecto a las evaluaciones que proponen. Otras se hallan en una fase más temprana de aprendizaje, en que siguen procurando identificar los casos en que convendría usar una evaluación de impacto. No obstante, siguen proponiéndose algunas evaluaciones de impacto en forma ad hoc tras el interés del especialista a cargo del proyecto. Ello apunta a la necesidad de un enfoque más estratégico en el Banco para definir las evaluaciones de impacto que han de realizarse.

Si bien el costo de las evaluaciones de impacto varía bastante, es ampliamente comparable al de otras organizaciones. En el BID, la mayor parte del costo de las evaluaciones de impacto recae sobre los clientes. El principal componente de costo de esas evaluaciones es la recopilación de datos, y en algunos casos esos datos pueden tener otros usos para el cliente. Sin embargo, la calidad de bien público que caracteriza a todos los productos de conocimiento podría dar lugar a una subinversión. Cuando no es necesario recopilar datos, el costo es mucho más bajo y la consideración principal es cómo conjugar los incentivos con los servicios acertadamente de modo de facilitar su financiamiento y ejecución.

El Banco ha avanzado hacia la elaboración de evaluaciones de mejor calidad, pero todavía sigue habiendo margen para mejoras. La calidad de las evaluaciones de impacto se juzgó en función de las normas internacionales y se consideró la pertinencia de las preguntas de evaluación, la idoneidad de los datos utilizados, el rigor de la metodología y la solidez del análisis. Alrededor del 55% de las evaluaciones completadas se calificaron como de calidad satisfactoria o parcialmente satisfactoria. Según el análisis realizado, si bien esa calidad disminuyó tras la adopción de la DEM debido a que otros sectores que normalmente no realizaban evaluaciones de impacto empezaron a prepararlas, luego el nivel siguió mejorando. Asimismo, la calidad de las propuestas también ha seguido una clara tendencia de mejora.

Pese a la gran campaña para producir evaluaciones de impacto, estas no se usan ampliamente en los documentos del BID, aunque pueden ser una importante fuente de información para discontinuar programas ineficaces. La mayor parte de los 25 clientes entrevistados para este informe indicaron que la inclusión de

evaluaciones de impacto en las operaciones era útil. Muchos plantearon que esas evaluaciones respaldaron la continuidad del programa objeto de evaluación y facilitaron la rendición de cuentas. Unos pocos señalaron que el Banco impone las evaluaciones de impacto y que se usan muy poco. OVE encontró pocas pruebas empíricas sobre el uso de las evaluaciones de impacto en las estrategias de país, las notas técnicas y los proyectos (a juzgar por las veces que se citan directamente), aunque las entrevistas señalan que desempeñan una función para fundamentar la decisión de discontinuar intervenciones o programas ineficaces.

Por último, el análisis brinda algunas pruebas que apuntan a que las evaluaciones de impacto podrían repercutir positivamente en la implementación de proyectos. Aunque el mecanismo causal no está claro, es posible que esto obedezca a mejoras en el diseño de los proyectos (que permiten que se realice una evaluación de impacto) o un mayor respaldo del equipo de evaluación en el terreno. Los proyectos del BID que incluyen evaluaciones de impacto desembolsan sus recursos con mayor rapidez y concluyen antes que los proyectos que no las incluyen, lo cual podría compensar hasta el 50% del costo de la evaluación.

Sobre la base de los hallazgos de la presente evaluación, OVE formula las siguientes recomendaciones para mejorar la producción y el uso de las evaluaciones de impacto en el BID:

1. Adoptar un enfoque más estratégico con respecto a la selección de las evaluaciones de impacto realizando o apoyando solo aquellas que cuenten con el compromiso del cliente (o una estrategia clara para lograr ese compromiso), una brecha de conocimiento bien definida y un cronograma factible. La presente evaluación demuestra la importancia de contar con el apoyo del cliente al realizar una evaluación de impacto. El compromiso del cliente también ayuda a asegurar que los resultados de las evaluaciones de impacto se empleen una vez concluidas.
2. Asegurar que los documentos de marco sectorial señalen las brechas de conocimiento para así ayudar a orientar la labor del Banco en materia de evaluaciones de impacto. A la fecha, los documentos de marco sectorial han hecho esto en distintas medidas, y OVE recomienda que de ahora en adelante lo hagan de forma más sistemática e integral.
3. Revisar la DEM a fin de reducir los incentivos que hacen que se proponga un número excesivo de evaluaciones de impacto. Todos los proyectos deben contar con un sistema sólido de seguimiento y evaluación, pero no todos deben incluir una evaluación de impacto. La DEM ha sido una herramienta útil para incrementar la evaluabilidad de los proyectos y promover una mayor

comprensión de la importancia de realizar evaluaciones más rigurosas. También ha incrementado los incentivos para proponer evaluaciones de impacto, lo que ha hecho que se generen propuestas que no estaban bien pensadas o que no se justificaban.

4. Crear un mecanismo de financiamiento transparente en el que los intereses de los clientes se correspondan con los del Banco. Cuando se considere pertinente y factible realizar una evaluación de impacto para una operación de préstamo, la evaluación debe incluirse en el sistema de seguimiento y evaluación que financie el préstamo en la medida en que resulte informativo para el país y para el informe de terminación de proyecto (ITP). Si hay efectos a más largo plazo que solo pueden medirse una vez se cierre el proyecto, debe financiarse una labor de evaluación ex post con recursos administrados por el Banco. De esa forma, los clientes comprometidos a realizar evaluaciones de impacto y a establecer mecanismos de seguimiento y evaluación adecuados podrán beneficiarse de recursos del BID para completar la evaluación después de que cierre el proyecto. Al mismo tiempo, el Banco comprometería los recursos propios que administra únicamente para las evaluaciones a más largo plazo en que haya un compromiso del cliente, una buena gestión inicial (con parámetros de referencia y seguimiento adecuados) y la expectativa de poder recabar información valiosa sobre los resultados del programa a más largo plazo.
5. Reforzar los sistemas de control de calidad. El Banco ya cuenta con una serie de mecanismos para el control de calidad ex ante. Sin embargo, conforme se muestra en la presente evaluación, no todas las evaluaciones de impacto se completan con un alto nivel de calidad. Para evitar el riesgo reputacional de publicar y difundir trabajos de baja calidad, el Banco necesita crear un sistema de control de calidad que vaya más allá de las evaluaciones de impacto que se publican como documentos de trabajo o notas técnicas.
6. Fortalecer y centralizar los sistemas del Banco para dar seguimiento a todas las evaluaciones de impacto que están en marcha —desde la fase de diseño hasta la final— y los costos y poner al alcance del público las evaluaciones de impacto de alta calidad y su respectiva documentación, los datos empleados (de ser posible) y las lecciones aprendidas. Como se señala en la presente evaluación, SPD y KNL han avanzado en este sentido, y el sistema ha de finalizarse.
7. Estructurar un mecanismo más formal para promover la formación de asociaciones que aumenten la calidad, la pertinencia y la visibilidad de la labor del BID, al tiempo que se reduzcan los costos. La presente

evaluación detectó pocos casos en que se crearon asociaciones para elaborar las evaluaciones de impacto en sí. Sobre la base de las buenas prácticas observadas en el BID y otros organismos, podría ser de utilidad que el Banco estableciera más asociaciones con círculos académicos e intensificara su colaboración con otros organismos en la producción y difusión de las evaluaciones de impacto y en el intercambio de información. SPD ha estado promoviendo esas actividades, y ese esfuerzo ha de fortalecerse.

- ¹ Cameron, Mishra y Brown (2016) crearon una inmensa base de datos de evaluaciones de impacto publicadas entre 1981 y 2012 para hacer un balance de las lecciones aprendidas y demostraron que si bien al inicio las evaluaciones se utilizaron principalmente en la literatura médica, desde principios de los años 2000 se han empleado cada vez más para evaluar políticas en numerosos países —con una cierta concentración en América Latina y Asia— basándose en una diversidad de métodos, no solo en los de tipo experimental.
- ² Center for Global Development (2006b).
- ³ What Have We Learned? Improving Development Policy Through Impact Evaluation. Seminario disponible en <https://www.povertyactionlab.org/event/what-have-we-learned-improving-development-policy-through-impact-evaluation>.
- ⁴ Véanse Cunha, Heckman y Schennach (2010), así como Bedoya et al. (2017).
- ⁵ Un préstamo del BID se utilizó para financiar PROGRESA en México. Levy (2004) y Levy y Rodríguez (2005) describen cómo se diseñó, ejecutó y evaluó el programa. Las autoridades mexicanas crearon el programa de transferencias condicionadas de efectivo a partir de información empírica recopilada en decenios previos sobre la dinámica de la asignación de los recursos de los hogares. Desconocían, sin embargo, los efectos específicos del programa, motivo por el que se diseñó una evaluación de impacto. Las pruebas empíricas recopiladas de la evaluación fueron fundamentales para reproducir el modelo en otros países y promover su continuidad en México, principalmente a través del respaldo del BID.
- ⁶ El Banco Interamericano de Desarrollo ha ejercido y sigue ejerciendo una función activa en el establecimiento de esos programas [de desarrollo humano focalizado] y en la tarea de asegurar que incluyan sistemas sólidos de evaluación de impacto (Legovini y Regalia, 2001).
- ⁷ Muy pocos proyectos definieron indicadores de desempeño coherentes y fueron muy pocos los que establecieron datos de referencia o que fijaron metas o hitos explícitos. En los casos en que existían metas e indicadores empíricos, solo se aplicaron a los productos del proyecto, no a los resultados que se preveía alcanzar en los países una vez generados esos productos (OVE, 2002).
- ⁸ Las evaluaciones de impacto de OVE normalmente se realizan ex post habida cuenta de su independencia de las operaciones cotidianas. Véase en Ruprah (2008) un análisis de los retos que plantea realizar evaluaciones de impacto ex post y de la forma en que OVE los abordó.
- ⁹ Entre 2006 y 2012, el porcentaje de funcionarios contratados que poseían un grado de maestría aumentó de 54% a 59%, en tanto que el porcentaje de especialistas contratados con grados de doctorado aumentó de 13% a 23% (OVE, 2014).
- ¹⁰ Véanse más detalles en el documento GA-232 del BID.
- ¹¹ El Marco de Efectividad en el Desarrollo es actualmente una Política Operativa del Banco.
- ¹² Véase el documento AB-3008, párrafo 4.23.
- ¹³ Los retos de financiar evaluaciones de impacto en los sectores público y privado son muy diferentes. Al prepararse la presente evaluación, la CII había completado una evaluación y tenía unas pocas programadas. Según la evaluación del FOMIN a cargo de OVE (documento MIF/RE-2), el FOMIN ha tomado medidas para fortalecer la preparación, el seguimiento y la evaluación de proyectos, y ha invertido en un programa de evaluaciones de impacto. Sin embargo, no alcanzó su meta de evaluar 25 proyectos por año.
- ¹⁴ Las evaluaciones de impacto pueden producirse (i) como productos de investigación pura, (ii) durante la preparación e implementación de los proyectos, (iii) como parte de trabajos sectoriales independientes o (iv) ex post como parte del ciclo regular de evaluación de proyectos del Banco.

- ¹⁵ Ello no necesariamente significa que no se realizó una evaluación. En la mayor parte de los proyectos que no fueron cancelados, la evaluación de impacto se reemplazó por otro tipo de análisis.
- ¹⁶ De las operaciones con evaluaciones de impacto, 105 están programadas para concluir en 2017 y un número cada vez menor en los años subsiguientes (98, 58, 28, 36 y 3). Sin embargo, normalmente las evaluaciones de impacto se finalizan algún tiempo después de concluida la operación.
- ¹⁷ A pesar del alto nivel de respuesta en general a las solicitudes de entrevista por parte de OVE, hubo un porcentaje más elevado de falta de respuesta entre los jefes de equipo de evaluaciones de impacto canceladas.
- ¹⁸ Se les preguntó a los jefes de equipo el motivo por el que habían propuesto cada una de las evaluaciones de impacto, concretamente, si se debía a interés del Banco o el cliente, al puntaje en la DEM u otras razones. Si la respuesta era la DEM, la probabilidad de cancelación era 10 puntos porcentuales más alta.
- ¹⁹ En alrededor del 18% de los ensayos aleatorios controlados, la estrategia de identificación se reemplazó por otros métodos.
- ²⁰ Esta evolución también puede observarse fuera del BID, pues se han desplegado mayores esfuerzos en sectores que no solían recurrir a evaluaciones de impacto (Cameron, Mishra y Brown, 2016).
- ²¹ Otros departamentos también han recurrido considerablemente a recursos de actividades económicas y sectoriales para financiar las evaluaciones de impacto, pero la mayoría se identificó por medio de operaciones de cooperación técnica incluidas en el análisis.
- ²² En los últimos decenios, los métodos para recopilar datos —censos y encuestas— se han centrado en indicadores sociales, lo cual ha facilitado el acceso a datos de calidad que pueden utilizarse para evaluar intervenciones en esos sectores.
- ²³ Véase información detallada en el Cuadro B.1 del Anexo II.
- ²⁴ Muchas de las cancelaciones fueron reemplazadas por un análisis económico ex post. Debido a la falta de uniformidad en las respuestas dadas en las entrevistas con OVE, resultó imposible identificar cada caso.
- ²⁵ Ello es especialmente esencial en sectores y países en que las encuestas nacionales o los registros administrativos son limitados. La información recopilada a través de la evaluación de impacto representa la única fuente de información para evaluar el impacto de la intervención. Un ejemplo de ello serían los programas de transferencias condicionadas de efectivo en Centroamérica.
- ²⁶ Gertler et al. (2016), Kaplan (2016), así como Karlan y Appel (2016).
- ²⁷ Este hallazgo se desprende de entrevistas a una muestra de clientes del BID que participaron en evaluaciones de impacto del BID.
- ²⁸ El porcentaje de consultores es más alto en INE e INT, donde los consultores realizan más del 85% de las evaluaciones. En esos casos la participación de las contrapartes en el diseño y la implementación sigue siendo muy limitada.
- ²⁹ La diferencia entre asociaciones y consultores está relacionada con la necesidad de pagar por el trabajo únicamente.
- ³⁰ Es más probable que las evaluaciones de impacto completadas hayan sido diseñadas en la primera parte del período objeto de análisis. En las entrevistas, muchos funcionarios explicaron que buscaron apoyo de expertos externos en las primeras operaciones que incluyeron una evaluación de impacto. Actualmente, recurren principalmente a recursos internos.

- ³¹ Los modelos de DIME y J-PAL, por ejemplo, demuestran que las asociaciones estratégicas crean condiciones que fomentan la creatividad y la especialización para desarrollar conocimiento nuevo. DIME aprovecha la red y el alcance del Banco Mundial; trabaja con jefes de equipo, representantes de países, gobiernos, especialistas de diferentes sectores y miembros de entidades internacionales especializadas para fortalecer las evaluaciones de impacto que produce y amplificar su uso y difusión. J-PAL emplea una red de académicos cuidadosamente seleccionada e integrada en las mejores universidades del mundo para intercambiar conocimiento y experiencia, establecer parámetros de referencia y mantener la elevada calidad reputacional de su labor.
- ³² Esos fondos se cobran al prestatario como parte de los préstamos. Normalmente no es posible separar los recursos asignados al seguimiento de los que se asignan a las evaluaciones de impacto durante la implementación del proyecto.
- ³³ Por ejemplo, los productos de conocimiento y fortalecimiento de capacidad (ahora heredados) fueron en gran medida financiados con recursos de operaciones de cooperación técnica.
- ³⁴ Todos los países tienen al menos una evaluación de impacto realizada con datos recopilados. Entre los países que tienen más de una se incluyen los siguientes: Costa Rica y Trinidad y Tobago (2 cada uno); Belize (4); Guatemala y Honduras (5 cada uno); Nicaragua (6); Jamaica (7); El Salvador (8); Haití y Paraguay (9 cada uno); y Bolivia (11).
- ³⁵ El Anexo III ofrece información más detallada sobre la forma en que estas organizaciones realizan las evaluaciones de impacto y los costos correspondientes.
- ³⁶ El monto del préstamo no incluye financiamiento no reembolsable para inversión.
- ³⁷ Esto no incluye operaciones de cooperación técnica reembolsable o financiamiento no reembolsable para inversión.
- ³⁸ Ello se ajusta a la estrategia que la CII está formulando para orientar sus evaluaciones de impacto.
- ³⁹ Ese proceso de revisión es la base en que se sustenta la revisión cualitativa de los trabajos académicos de todas las publicaciones académicas. Asimismo, es la base empleada en la selección de todos los trabajos de investigación financiados por las principales entidades de financiamiento, públicas y privadas, en todo el mundo.
- ⁴⁰ CLEAR-LAC se seleccionó a través de una licitación pública. Un equipo de seis investigadores aplicó el instrumento de revisión de la calidad. La mayoría de los participantes eran investigadores del Centro de Investigación y Docencia Económicas (CIDE), de formación económica y con estudios en ciencias cuantitativas y experiencia previa en la ejecución o revisión de evaluaciones de impacto. La distribución de las evaluaciones de impacto entre los participantes se realizó teniendo en cuenta la pericia sectorial de cada investigador. Un subconjunto de cuestionarios de evaluación fue objeto de una revisión de pares a fin de aclarar los criterios necesarios para completar el instrumento.
- ⁴¹ El marco utilizado para este propósito se emplea comúnmente en literatura especializada para identificar las características básicas de las evaluaciones de impacto que proporcionan pruebas empíricas convincentes para sustentar lo que funciona y lo que no funciona en diferentes disciplinas; véase Sherman et al. (1998), Halpern (2014), Gertler et al. (2016) y Karlan y Appel (2016).
- ⁴² Esta muestra abarca todas las evaluaciones completadas, con la excepción de siete documentos de muy baja calidad que recibió OVE todavía en forma de borradores, y uno publicado en *Quarterly Journal of Economics*, que fue objeto de un riguroso control de calidad.

- ⁴³ Se trata de una muestra estratificada por departamento.
- ⁴⁴ Se seleccionó una muestra estratificada aleatoria de 59 propuestas de evaluaciones de impacto para realizar el análisis del examen de calidad. CLEAR-LAC llevó a cabo el mismo análisis empleando otra versión del marco de calidad que elaboró OVE.
- ⁴⁵ Es probable que el número de personas que asistieron a esas actividades de capacitación sea menor pues una misma persona puede participar en varias actividades.
- ⁴⁶ Véase más información en el Anexo VIII.
- ⁴⁷ Por medio del programa informático MAXQDA se hizo un análisis léxico de los documentos del BID, incluidos documentos de aprobación de préstamos, estrategias de país y notas técnicas, para encontrar el término “evaluación de impacto” (en inglés y español). Las limitaciones de dicho análisis obedecen a la naturaleza de su objetivo, pues es bien sabido que las fuentes académicas no se citan debidamente en las decisiones políticas, los diseños de intervenciones y los documentos de proyecto (Behrman, 2010).
- ⁴⁸ Banco Mundial (2009).
- ⁴⁹ En el 30% de los casos el Banco ya no tenía contacto con representantes del gobierno. Otros factores fueron cambios recientes de gobierno, pérdida de memoria institucional acerca de la evaluación o lo inoportuno del momento, a juicio del especialista del Banco, para contactar a la contraparte.
- ⁵⁰ En más de la mitad de los casos, el Banco parece haber hecho un esfuerzo especial por capacitar al organismo ejecutor.
- ⁵¹ Los tipos de publicaciones en donde se incluyen las evaluaciones de impacto del BID son principalmente las notas técnicas (36%); también se divulga un porcentaje considerable en publicaciones especializadas revisadas por pares (13%), a través de portales de investigación como IDEAS y SSRN (13%) y en sitios virtuales de contrapartes gubernamentales (8%).
- ⁵² La colección, que se recopiló utilizando el programa Publish or Perish que creó Anne-Wil Harzing de la Universidad de Melbourne, incluye campos de búsqueda para indicar la afiliación del autor, el año de publicación y términos clave. La búsqueda incluyó toda referencia al Banco (“BID”, “Banco Interamericano de Desarrollo” o “IADB”), la expresión “evaluación de impacto” y los años de publicación entre 2006 y 2016. El programa informático se basa en Google Scholar y calcula el número de citas por artículo, así como parámetros generales de la búsqueda de citas. Aunque se asemeja a las bases de datos SCOPUS y Thompson ISI de uso común, su cobertura de publicaciones en idiomas distintos al inglés es más amplia e incluye documentos de trabajo y libros, así como artículos académicos de publicaciones especializadas revisadas por pares. Se llevaron a cabo búsquedas análogas en los cuatro idiomas oficiales del BID.
- ⁵³ El análisis de citas presentado en este informe se realizó en octubre de 2016 y, por lo tanto, abarca datos hasta septiembre de 2016.
- ⁵⁴ Este análisis se llevó a cabo utilizando bases de datos que mantienen SCOPUS y Thompson-ISI. Para años más recientes se realizaron análisis similares con el programa informático Publish or Perish que se emplea como interfaz para recopilar los datos sobre citas del programa Microsoft Academic. Se llega a la misma conclusión cuando se comparan el BID, el Banco Mundial, el Banco Africano de Desarrollo y el Banco Asiático de Desarrollo.
- ⁵⁵ En el Anexo VII se presentan detalles sobre el método econométrico empleado y el análisis completo.

- ⁵⁶ Este análisis, que es solo de carácter indicativo, proporciona una estimación general del ahorro generado por las tasas de desembolso promedio de los préstamos del BID. Ese ahorro puede variar significativamente por sector, vencimiento y país.
- ⁵⁷ El monto promedio de un préstamo para inversión es de US\$83,9 millones si incluye una evaluación de impacto y de US\$92,7 millones si no la incluye. La comisión de crédito asciende actualmente a 0,5%, y en este ejemplo se aplicó su equivalente mensual. Cuando el monto utilizado en la comparación es el mismo en ambos casos, el ahorro varía entre 14% (US\$50 millones) y 25% (US\$100 millones).

- Banco Mundial (2009). *Making Smart Policy: Using Impact Evaluation for Policy Making— Case Studies on Evaluations that Influenced Policy*. Doing Impact Evaluation #14. Washington, D.C.: Banco Mundial.
- Banerjee, Abhijit y Esther Duflo (2008). *The Experimental Approach to Development Economics*. Department of Economics y Abdul Latif Jameel Poverty Action Lab. Cambridge, MA: Instituto de Tecnología de Massachusetts.
- Bedoya, Guadalupe, Lucca Bittarello, Jonathan Davis y Nikolas Mittag (2017). *Distributional Impact Analysis. Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect*. Policy Research Working Paper 8139. Washington, D.C.: Banco Mundial.
- Behrman, J. R. (2010). *The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash Program*. World Development, 38(10): 1473–1485.
- Blattman, Chris (2008). *Impact Evaluation 2.0*. Presentación al Ministerio Británico para el Desarrollo Internacional. Londres, 14 de febrero de 2008: DFID.
- Cameron, D. B., A. Mishra y A.N. Brown (2016). *The growth of impact evaluation for international development: How much have we learnt?* Journal of Development Effectiveness 8(1): 1-21.
- Center for Global Development (2006a). *A Major Step Forward on Impact Evaluation*. Disponible en: <http://www.cgdev.org/content/article/detail/8207/>.
- _____ (2006b). *When Will We Ever Learn? Improving Lives Through Impact Evaluation*. Informe de trabajo de la brecha de evaluación. Washington, D.C.: Center for Global Development.
- Cunha, F., J. Heckman y S. Schennach (2010). *Estimating the Technology of Cognitive and Non-Cognitive Skill Formation*. Econometrica 78(3): 883-931.
- Deaton, Angus (2009). *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*. NBER Working Paper No. 14690.
- Deaton, Angus (2010). *Instruments, Randomization, and Learning about Development*. Journal of Economic Literature, vol. 48 (2): 424-55.
- Dhaliwal, I. y C. Tulloch (2012). *From Research to Policy: Using Evidence of Impact Evaluations to Inform Development Policy*. Journal of Development Effectiveness 4(4): 515-536.
- Duflo, E. y M. Kremer (2003). *Use of Randomization in the Evaluation of Development Effectiveness*. Informe preparado para la Conferencia sobre Evaluación y Efectividad en el Desarrollo, 15 y 16 de julio, Banco Mundial, Washington, D.C.

- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings y Christel M. J. Vermeersch (2016). *Impact Evaluation in Practice*. Segunda edición. Washington, D.C.: Banco Interamericano de Desarrollo y Banco Mundial.
- Halpern, David (2014). *What Works? Evidence for Decision Makers*. Londres: Ministerio Británico para el Desarrollo Internacional.
- Imbens, Guido W. y Donald B. Rubin (2008). *Rubin Causal Model*. En: The New Palgrave Dictionary of Economics, segunda edición, compilación a cargo de Steven N. Durlauf y Lawrence E. Blume. Nueva York: Palgrave.
- International Initiative for Impact Evaluation (2011). *3IE Impact Evaluation Practice: A Guide for Grantees*. Disponible en: www.3ieimpact.org/strategy/pdfs/3ie_impact_evaluation_practice.pdf.
- Karlan, Dean y Jacob Appel (2011). *More than Good Intentions*. Nueva York: Dutton Press.
- Karlan, Dean y Jacob Appel (2016). *Failing in the Field: What We Can Learn When Field Research Goes Wrong*. Princeton: Princeton University Press.
- Legovini, Arianna y Ferdinando, Regalia (2001) *Targeted Human Development Programs: Investing in the Next Generation*. Departamento de Desarrollo Sostenible, serie de practicas óptimas. Washington, D.C.: Banco Interamericano de Desarrollo.
- Legovini, Arianna, Vincenzo Di Maro y Caio Piza (2015). *Impact evaluation helps deliver development projects*. Policy Research Working Paper no. WPS 7157; Impact Evaluation series. Washington, D.C.: Grupo Banco Mundial.
- Levine, Ruth (2010). Palabras ante la Tercera Conferencia Anual sobre la Evaluación de Programas en el Departamento de Estado (Washington, D.C.). 9 de junio de 2010.
- Levy S. y E. Rodriguez (2004). *Economic crisis, political transition, and poverty policy reform: Mexico's PROGRESA-Oportunidades Program*. Serie de diálogo de políticas. Washington, D.C. BID.
- Levy, Santiago y Evelyne Rodríguez (2005). Sin Herencia de Pobreza: El Programa Progres-Oportunidades de México. Washington, D.C: Banco Interamericano de Desarrollo.
- Ludwig, Jens, Jeffrey R. Kling y Sendhil Mullainathan (2011). *Mechanism Experiments and Policy Evaluations*. Journal of Economic Perspectives 25 (3): 17–38.
- OCDE. (2006). *Outline of Principles of Impact Evaluation*. Disponible en: <http://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf>
- Oficina de Evaluación y Supervisión (2002). Análisis de la Evaluabilidad de los

- Proyectos, 2001 (documento RE 275). Washington, D.C.: Banco Interamericano de Desarrollo.
- _____ (2014). Evaluación de los Resultados de la Realineación (documento RE 451-2). Washington, D.C.: Banco Interamericano de Desarrollo.
- Pritchett, L. (2002). *It pays to be ignorant: A simple political economy of rigorous program evaluation*. The Journal of Policy Reform 5(4): 251-269.
- Pritchett, Lant y Justin Sandefur (2013). *Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix*. Center for Global Development Working Paper 336.
- Ravallion, M. (2008). *Evaluation in the Practice of Development*. Policy Research Working Paper 4547, Development Research Group. Washington, D.C.: Banco Mundial.
- Rothstein, A., J. Sutton y M. Borenstein (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, Inglaterra y Hoboken, NJ: Wiley.
- Ruprah, I. J. (2008). *You Can Get It If You Really Want: Impact Evaluation Experience of the Office of Evaluation and Oversight of the Inter-American Development Bank*. Boletín del BID Número 39(1): páginas 23-35.
- Savedoff, W. D., R. Levine y N. Birdsall (2006). *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Washington, D.C.: Center for Global Development. Disponible en <http://www.cgdev.org/content/publications/detail/7973>.
- Sherman, Lawrence W.; Denise C. Gottfredson, Doris L. MacKenzie, John Eck, Peter Reuter y Shawn D. Bushway (1998). *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, D.C.: Instituto Nacional de Justicia.
- Sutcliffe, Sophie y J. Court (2005): *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?* Documento de trabajo del Instituto de Desarrollo de Ultramar (ODI).
- The Network of Networks on Impact Evaluation (2009) Impact Evaluations and Development. NONIE Guidance on Impact Evaluation*. Grupo de Evaluación Independiente. Banco Mundial, Washington, D.C.
- Vivalt (2016). *How Much Can We Generalize from Impact Evaluations?* Informe inédito.
- Weyrauch, V. y G. Díaz Langou (2013). *Sound expectations: From impact evaluations to policy change*. Journal of Development Effectiveness 5(3): 269-304.