



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE DOCTORADO EN GEOGRAFÍA

Centro de Investigaciones en Geografía Ambiental

Comparación interorganizacional del uso de evaluaciones en una reserva de la biosfera

TESIS QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN GEOGRAFÍA

PRESENTA:

EDUARDO LUNA SÁNCHEZ

DIRECTORA DE TESIS

DRA. MARGARET SKUTSCH

Centro de Investigaciones en Geografía Ambiental - UNAM

COMITÉ TUTOR

DRA TUYENI HEITA MWAMPAMBA

Instituto de Investigaciones en Ecosistemas y Sustentabilidad - UNAM

DR. RAÚL FRANCISCO PINEDA LÓPEZ

Universidad Autónoma de Querétaro

MORELIA, MICHOACÁN, FEBRERO 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Comparación interorganizacional del uso de evaluaciones en una reserva de la biosfera

Fotografía: Eduardo Luna

EL HOMBRE EN LA ARENA

No es el crítico quien cuenta, ni el que señala con el dedo al hombre fuerte cuando tropieza, o el que indica en qué cuestiones quien hace las cosas podría haberlas hecho mejor. El mérito recae exclusivamente en el hombre que se halla en la arena, aquel cuyo rostro está manchado de polvo, sudor y sangre, el que lucha con valentía, el que se equivoca y falla el golpe una y otra vez, porque no hay esfuerzo sin error y sin limitaciones.

El que cuenta es el que de hecho lucha por llevar a cabo las acciones, el que conoce los grandes entusiasmos, las grandes devociones, el que agota sus fuerzas en defensa de una causa noble, el que, si tiene suerte, saborea el triunfo de los grandes logros y si no la tiene y falla, fracasa al menos atreviéndose al mayor riesgo, de modo que nunca ocupará el lugar reservado a esas almas frías y tímidas que ignoran tanto la victoria como la derrota.

-Theodore Roosevelt-

Agradecimientos

A Mayra, que con tus manos de cuchara has tomado mis dedos chuecos para escribir las palabras con las que los espíritus cantan por las noches: "Fuego, camina conmigo".

A mi papá, Jesús. Eres un padre con el que se puede viajar en un burro negro toda la noche, por terrenos baldíos, hasta el amanecer, sin extraviarse jamás.

A mi mamá, María Cristina. Más allá de máscaras y figuras marinas, moldeas un amor cercano y creativo que me hace sentir en casa aún a la distancia.

A mi carnal, Álvaro. Te quiero tanto que te recibiría con entusiasmo y con lágrimas de alegría así vinieras del infierno.

A Margaret Skutsch. Gracias por la generosidad, calidez y confianza que caracterizaron tu acompañamiento a lo largo de este proceso formativo.

Índice

RESUMEN.....	1
INTRODUCCIÓN	3
CAPÍTULO 1. ENMARCANDO EL PROBLEMA DEL USO DE EVALUACIONES EN EL CAMPO DE LA CONSERVACIÓN.....	14
1.1 CONCEPTOS BÁSICOS DE EVALUACIÓN.....	15
1.2. LAS EVALUACIONES COMO PROCESO	39
1.3 SÍNTESIS DE CONCEPTOS CLAVE	66
CAPÍTULO 2. ABORDAJE METODOLÓGICO PARA ESTUDIAR EL USO DE EVALUACIONES	70
2.1 ESTRATEGIAS PARA ESTUDIAR EL USO DE EVALUACIONES	71
2.2 DISEÑO DE LA INVESTIGACIÓN.....	75
CAPÍTULO 3. CONTEXTO DEL ESTUDIO DE CASO	90
3.1 REGIÓN CENTRO Y EJE NEOVOLCÁNICO.....	91
3.2 CARACTERÍSTICAS GENERALES DE LA RESERVA DE LA BIOSFERA.....	94
3.3 ORGANIZACIONES ESTUDIADAS	95
CAPÍTULO 4. ¿SIRVEN LAS EVALUACIONES PARA APRENDER? INFLUENCIA DE LAS EVALUACIONES SOBRE LAS DECISIONES DE MANEJO EN UN ÁREA NATURAL PROTEGIDA.....	108
CAPÍTULO 5. EL USO DE EVALUACIONES EN UN CENTRO PÚBLICO DE INVESTIGACIÓN AGRÍCOLA	130
CAPÍTULO 6. ¿MEDIMOS LO QUE ES VALIOSO O VALORAMOS LO QUE ES MEDIBLE? COMPARACIÓN INTERORGANIZACIONAL DEL USO DE EVALUACIONES EN UNA RESERVA DE LA BIOSFERA.....	152
CAPÍTULO 7. ¿QUÉ NECESITA CAMBIAR PARA PERMITIR EL APRENDIZAJE A PARTIR DE EVALUACIONES?	174
7.1 CASO DE ESTUDIO 1- CONANP	176
7.2. CASO DE ESTUDIO 2- UNIVERSIDAD PÚBLICA	180
7.3 COMPARACIÓN INTERORGANIZACIONAL DEL USO DE EVALUACIONES	184
7.4 REFLEXIONES FINALES	194
REFERENCIAS.....	205
ANEXOS	227

Índice de mapas, cuadros y figuras

Mapas

Mapa 1. Reservas de la Biosfera ubicadas dentro de la Dirección Regional Centro y Eje Neovolcánico	92
--	----

Cuadros

Cuadro 1. Diferentes formas de abordar la evaluación de AP.....	6
Cuadro 2. Tres escenarios donde la evaluación prueba las proposiciones de un proyecto.....	16
Cuadro 3. Terminología básica útil para establecer relaciones de causalidad.....	17
Cuadro 4. Principios para la integridad metodológica de evaluaciones de impacto.....	21
Cuadro 5. Jerarquía de fuentes de información de conocimiento seguro sobre las relaciones de causa-efecto en las intervenciones.....	21
Cuadro 6. Diferencia entre outcomes e impactos en la inferencia contrafactual.....	22
Cuadro 7. Influencia del contexto sobre la efectividad del mecanismo.....	23
Cuadro 8. Ejemplos de relación principal-agente en el campo del desarrollo.....	29
Cuadro 9. Diferencias entre los dos tipos de evaluaciones sumativas.....	30
Cuadro 10. Algunas herramientas útiles para promover el aprendizaje organizacional en el campo del desarrollo.....	35
Cuadro 11. Comparación de evaluaciones sumativas y formativas.....	36
Cuadro 12. Transición hacia una rendición de cuentas transformacional.....	38
Cuadro 13. Estándares internacionales de la evaluación de programas.....	40
Cuadro 14. Descripción de cinco tipos de uso de evaluaciones.....	41
Cuadro 15. Contribuciones de la teoría del cambio (TdC) a la evaluación de proyectos.....	47
Cuadro 16. Respuestas organizacionales ante el error.....	54
Cuadro 17. Incentivos intangibles y beneficios no económicos para el manejo basado en resultados.....	56
Cuadro 18. Factores que inciden en la producción y uso de evaluaciones.....	67
Cuadro 19. Tipos de evaluación.....	68
Cuadro 20. Usos de la evaluación.....	68
Cuadro 21. Diferentes formas de estudiar el uso de evaluaciones en el campo del desarrollo.....	73
Cuadro 22. Relación de organizaciones seleccionadas en comparación con el total de organizaciones similares con presencia en la reserva.....	82
Cuadro 23. Marco para valorar si es posible interactuar con las organizaciones para promover cambios en su sistema de evaluación.....	84
Cuadro 24. Número de entrevistas y de documentos revisados.....	86
Cuadro 25. Ejes sustantivos de la estrategia 2040 de la CONANP.....	96
Cuadro 26. Algunos indicadores de objetivo establecidos en el Programa Nacional de Áreas Naturales Protegidas.....	98
Cuadro 27 Algunos indicadores del Plan Institucional de Desarrollo 2015-2018.....	101
Cuadro 28. Objetivos y líneas de acción ligados al manejo de recursos naturales en el plan de desarrollo municipal 2015-2018.....	103
Cuadro 29. Jerarquización de las organizaciones en función del potencial de sus evaluaciones como herramienta de aprendizaje.....	185
Cuadro 30. Factores que promueven o inhiben el uso de evaluaciones como una herramienta de aprendizaje en las organizaciones estudiadas.....	186
Cuadro 31. Diferencias entre victorias rápidas y pequeñas.....	197
Cuadro 32 Formas concretas de conceptualizar las victorias pequeñas.....	198
Cuadro 33. Respuestas para diferentes tipos de falla.....	200
Cuadro 34. Plan de colaboración con la universidad pública estudiada.....	202

Figuras

Figura 1. Cadena de resultados de un proyecto de conservación	19
Figura 2. Representación de la lógica de evaluaciones formativas.....	31
Figura 3. Niveles del aprendizaje organizacional.....	33
Figura 4. Factores que afectan la producción y uso de las evaluaciones.....	45
Figura 5. Creando el motivo, la oportunidad y los medios para el aprendizaje.	53
Figura 6. Influencia de los arreglos institucionales sobre el comportamiento organizacional	57
Figura 7. Vinculación causal de los conceptos clave empleados en la investigación.....	66
Figura 8. Secuencia lógica de la investigación.....	76
Figura 9. Técnicas de recolección de la información.....	85
Figura 10. Especies representativas y ecosistemas de la región Centro y Eje Neovolcánico.....	93
Figura 11. Capturas de pantalla de la página web del foro.....	233
Figura 12. Niveles de intervención para cambiar una cultura de investigación.	238

Imágenes

Imagen 1. Un equipo compartiendo sus reflexiones durante la actividad del café mundial	230
Imagen 2. Un equipo conversando durante la actividad del café mundial.....	232

Resumen

En el campo del desarrollo, la razón de ser de la evaluación es mejorar la eficacia y eficiencia de las organizaciones, incrementando así su capacidad de conseguir los objetivos que persiguen. Dada su vocación práctica, se espera que la información proveniente de ellas genere consecuencias, es decir, que se use para tomar decisiones y manejar los proyectos a través de adaptaciones. Pero ¿en verdad sirven las evaluaciones para aprender?, ¿miden lo que se necesita conocer para mejorar el desempeño? El objetivo de la investigación fue comparar la forma en que distintas organizaciones hacen y usan evaluaciones en una reserva de la biosfera del centro de México, con la intención de establecer similitudes y diferencias en el grado que utilizan evaluaciones como herramienta de aprendizaje organizacional. El fenómeno se describió en cuatro organizaciones que gestionan proyectos vinculados con el manejo de recursos naturales: una administración municipal, una dependencia del gobierno federal, una universidad pública y una Organización de la Sociedad Civil (OSC). Los resultados obtenidos muestran diferencias importantes entre ellas. Mientras que la dependencia federal y la OSC tienen los sistemas de evaluación más diversificados, cubriendo propósitos ideales de rendición de cuentas, aprendizaje organizacional y gestión de financiamiento, el sistema de evaluación empleado por la universidad pública es más limitado y en el caso de la administración municipal inexistente. Por ello, a nivel de los artefactos *per se*, la dependencia federal y la OSC están mejor posicionadas para aprender a partir de evaluaciones. A su vez, se resaltaron una serie de factores que dan forma a los sistemas de evaluación de las organizaciones, los cuales, explican por qué existen desconexiones entre la teoría y la práctica de sus evaluaciones. Estos se sistematizaron distinguiendo tres niveles, el metodológico, el organizacional y el institucional. Si bien existen excepciones, el personal de las organizaciones refirió la predominancia de evaluaciones sumativas basadas en la entrega de resultados cuantificables predeterminados. En consecuencia, un margen limitado para reportar fallas y resultados inesperados, y un sesgo hacia el reporte de números por sobre análisis de procesos, son obstáculos importantes para el uso de evaluaciones formativas.

Abstract

In the development field, the *raison d'être* of evaluation is to improve the effectiveness and efficiency of organizations so that they are able to deliver better on their central goals. It is expected that the information derived from evaluations will generate consequences, specifically to make decisions and manage projects through adaptations. But are evaluations really useful for learning? Do they measure what is needed to improve performance? The objective of the research was to compare the way in which different organizations used evaluations in a biosphere reserve located in central Mexico, with the intention of establishing similarities and differences in the degree to which they used evaluations as a learning tool. The phenomenon was described in four organizations that implement natural resource management projects: a local government, a federal agency, a public university, and a Non-Governmental Organization (NGO). The results show important differences between them. While the federal agency and the NGO have the most diversified evaluation systems, covering ideal purposes of accountability, organizational learning and fundraising, the evaluation system used by the public university is more limited and in the case of the local government non-existent. Thus, at the artifact level, both the federal agency and the NGO are better positioned to learn from evaluations. In turn, a series of factors that shaped the organizational evaluation systems were highlighted, explaining why there were disconnections between theory and practice. These were systematized, distinguishing between three levels: methodological, organizational and institutional. Although there were exceptions, the staff of the organizations mentioned the predominance of summative evaluations based on the delivery of predetermined quantifiable results. Consequently, the limited margin to report failures and unexpected results, and a bias towards reporting numbers over process analysis, are major obstacles to the use of formative evaluations.

INTRODUCCIÓN

Múltiples análisis del impacto de las decisiones humanas sobre la biosfera han concluido que atender simultáneamente la pérdida de biodiversidad, el cambio climático y la creciente inequidad de nuestra sociedad requiere transformar el sistema económico mundial (Turnhout *et al.*, 2021). Si bien, ese sistema ha sido capaz de cuatuplicar en 50 años el Producto Interno Bruto Mundial, su obsesión por el crecimiento ha implicado destrucción ambiental (Shukla *et al.*, 2019) y una distribución desigual de beneficios y costos entre la población mundial (UNDP, 2020).

Por ello, más allá de narrativas que enmarcan el problema en términos de la demanda de recursos asociada al crecimiento poblacional, se necesitan estrategias coherentes que, guiadas por valores de economía social, hagan énfasis en las conexiones entre comunidades, ecosistemas y justicia (Nagendra, 2018). Lo anterior implica pensar en alternativas de manejo integrado del paisaje, las cuales sean aplicables en porciones cada vez mayores de la tierra (Shukla *et al.*, 2019). Las áreas naturales protegidas (desde ahora ANP) son una de las opciones de política que se han propuesto para ese fin.

El propósito de las ANP es reducir presiones humanas en sitios que albergan diversidad biológica y que generan servicios ecosistémicos (Craigie *et al.*, 2015; Ferraro y Pressey, 2015). A nivel mundial, su valoración se refleja en la Meta 11 de Aichi, la cual establece el compromiso de conservar, para el año 2020, el 17% de las áreas terrestres y el 10% de las áreas marinas a través de ANP.

Existen diferentes visiones en torno del manejo de ANP. En un extremo se encuentra el modelo de fortalezas de la conservación, y en el otro, el de reservas de la biosfera. El primero asume que se obtienen mejores resultados en la protección de la biodiversidad si se garantiza el funcionamiento de los ecosistemas sin ningún tipo de alteración humana (Doolittle, 2007). El segundo parte del supuesto de que las personas y la naturaleza pueden coexistir y prosperar en el mismo espacio y a través del tiempo (Stairs, 2007).

En este sentido, las reservas de la biosfera son un modelo integrado de conservación y desarrollo que procura simultáneamente la protección ecológica y el bienestar de los pobladores locales (UNESCO, 2013). En estas la pregunta no es si los entornos naturales deben o no ser modificados, “sino las preguntas, mucho más complejas, de cantidad y calidad: a qué grado y de qué manera deberían modificarse los sistemas naturales para satisfacer necesidades humanas” (Perreault, 2009: 452). Las reservas reúnen una mezcla de políticas, programas e instrumentos que buscan responder esas preguntas. Ejemplo de ello son mecanismos de conservación basados en el mercado (ecoturismo, programas de pago por servicios ambientales) o iniciativas para mejorar la producción agropecuaria y el manejo forestal (pastoreo racional, agroforestería, o aprovechamiento de productos no maderables). Las intervenciones, no siempre

coordinadas entre sí, son implementadas por distintas organizaciones que interactúan con múltiples actores locales (Baird, 2014).

El modelo general de una reserva de biosfera distingue tres zonas. La primera es la zona núcleo, el área mejor conservada y que contiene la mayor biodiversidad. En estas áreas no se permiten actividades humanas que interfieran con la dinámica de los ecosistemas. Alrededor de ellas se definen zonas de amortiguamiento (*buffer zones* en inglés), las cuales permiten actividades humanas de bajo impacto, pero que garanticen la protección de las zonas núcleo. A su vez, las zonas de amortiguamiento son rodeadas por zonas de transición, en las cuales son válidos usos del suelo más intensivos, aunque congruentes con las metas de conservación (Guevara, 2010).

Actualmente, a nivel internacional existen 701 reservas de la biosfera distribuidas en 124 países (UNESCO, 2019). En México, las reservas de la biosfera son la categoría de conservación con mayor superficie protegida, 62 millones 952 mil 750 hectáreas distribuidas en 44 sitios a nivel nacional (Secretaría de Medio Ambiente y Recursos Naturales, 2018). Esto tiene sentido, dada la gran proporción del territorio forestal que es propiedad de comunidades indígenas y campesinas (Boege, 2008; Merino y Ortiz, 2013).

Pero, a pesar de su popularidad, hay que respaldar con evidencia si están teniendo éxito. Eso implica responder, para cada caso, el grado en el que están generando los resultados sociales¹ y ambientales² que se esperan (Ferraro y Pressey, 2015). Además, hay que comprender la efectividad de cada uno de los programas e instrumentos anidados en las reservas, los cuáles plantean múltiples actividades (capacitaciones, desarrollo de instrumentos normativos, manejo directo de hábitats) a diferentes escalas (ejidos, zonas núcleo, ecosistemas).

En el manejo de recursos naturales, es difícil cumplir múltiples objetivos con una sola herramienta (Andam *et al.*, 2008; Alix *et al.*, 2014; Clements *et al.*, 2014). En términos generales, se ha señalado el desconocimiento de los efectos que tienen las iniciativas sobre indicadores socioeconómicos y ambientales, por lo que se requiere información para trascender las decisiones basadas en supuestos y experiencias personales (Geldmann *et al.*, 2013; Roe *et al.*, 2013; Baylis *et al.*, 2016; McKinnon *et al.*, 2016).

A nivel internacional, el desconocimiento de la efectividad del desarrollo ha impulsado un interés en evaluaciones. Existen dos razones para esto. La primera es que se les considera un mecanismo para llamar a cuentas a las organizaciones que reciben recursos en nombre de la generación de valor público. Actualmente, se espera que las organizaciones demuestren que las intervenciones están bien

¹ P. ej. mejoras en la riqueza relativa, en la distribución equitativa de beneficios, o en el conocimiento del medio ambiente (Pomeroy *et al.*, 2004).

² P. ej. mejoras en la abundancia de especies clave, en la distribución del hábitat o en la integridad de la cadena trófica (Pomeroy *et al.*, 2004).

planificadas y que los recursos se usan de forma efectiva para conseguir resultados (Chaplowe y Cousins 2016: 14).

En torno de ANP, estas demandas se ven reflejadas en la meta 11 de Aichi, la cual, no solo solicita una expansión de la superficie bajo un estatus de protección, sino evidencia de que se maneja de forma efectiva (Hockings *et al.*, 2015). Además, en el financiamiento de proyectos, cada vez más donantes piden pruebas de que su dinero ha sido bien invertido en el cumplimiento de objetivos de conservación (Bertzky, 2009).

La segunda razón es la confianza en que las evaluaciones servirán para tomar decisiones inteligentes. Es decir, decisiones basadas en el análisis sistemático de información relevante para planificar una intervención o para concluir si está funcionando (Chelimsky, 1997).

Como herramienta de planificación, las evaluaciones mejoran el diseño de los proyectos al anticipar las hipótesis a probar y al pensar en los indicadores con los que se establecerá su valor de verdad. Además, de forma *ex ante*³, permiten prever la generación de impactos negativos a ser evitados o mitigados mediante acciones especiales (Richards, 2012).

Por su parte, como herramienta de análisis *ex post*⁴, la función de las evaluaciones es recolectar información para confirmar o refutar las hipótesis sobre las que se fundamentó la lógica de los proyectos. Por ejemplo, en un proyecto de conservación de jaguar, las evaluaciones sirven para validar el vínculo causal entre acciones (como talleres de sensibilización con ganaderos), resultados inmediatos (por ejemplo, adaptaciones en el sistema de producción de los ganaderos para disminuir eventos de depredación), efectos a generar (una disminución en los eventos de depredación y en la cacería de represalia) e impactos (poblaciones sanas de jaguar).

Sin embargo, dado que las intervenciones ocurren en contextos complejos⁵, una de las mayores expectativas es que las evaluaciones servirán para manejarlas a través de adaptaciones (Wild y Ramalingam, 2018; Hernandez *et al.*, 2019; Pasanen *et al.*, 2019). El concepto de 'manejo adaptativo' implica tratar, de forma deliberada, a las políticas, programas y proyectos como un experimento que pone a prueba sus hipótesis. El supuesto básico es que si durante la implementación se genera y usa evidencia para orientar la acción futura, entonces se incrementará la

³ Las evaluaciones *ex ante* o prospectivas cubren dos propósitos diferentes. En primer lugar, sirven para corregir supuestos equivocados en la lógica de una intervención (Peersman *et al.*, 2015). En segundo lugar, una vez establecido que es plausible conseguir los cambios deseados, las evaluaciones prospectivas permiten estimar los impactos potenciales para planear el manejo de las posibles consecuencias, tanto positivas como negativas, de las intervenciones planificadas (Esteves *et al.*, 2012).

⁴ Este tipo de evaluaciones establecen si un proyecto consiguió los resultados esperados. Su objetivo es rendir cuentas sobre el valor público generado, y a la vez, explicar cómo funcionó el proyecto y para quién (Barnett y Munslow, 2014; Van Hemelrijck y Guijt 2016).

⁵ Los escenarios complejos presentan valores e intereses en disputa y una red de problemas interrelacionados. Esto hace que la dinámica de los procesos de cambio sea incierta, por lo que se considera necesario aprender sobre la marcha para desempeñarse mejor (Guijt, 2008).

efectividad (Salafsky *et al.*, 2019; Pasanen *et al.*, 2019). Por ello, es crucial que las organizaciones brinden oportunidades para que sus integrantes aprendan, de forma que sean capaces de responder mejor al contexto cambiante que enfrentan (Argyris y Schön, 1978). Este es el propósito central de muchas metodologías de evaluación en el campo de la conservación y el desarrollo (Pomeroy *et al.*, 2004; Conservation Measures Partnership, 2007; Stolton y Dudley, 2016; Hockings *et al.*, 2006).

Ante estas aspiraciones, surgen las siguientes dudas fundamentales: ¿podemos confiar en que las evaluaciones mejoran los proyectos a través de los que se manejan las ANP? ¿Gracias a ellas se obtienen mejores resultados de conservación y desarrollo? El interés por estas preguntas impulsó el proyecto de investigación doctoral que se presenta en esta tesis.

La inquietud, o mejor, el malestar generado por el tema, se originó a partir de mis experiencias como evaluador externo de programas ambientales de la Comisión Nacional Forestal y de una serie de Organizaciones de la Sociedad Civil. En ellas, detecté una distancia preocupante entre el producto de una evaluación y su uso para los fines previstos por las metodologías. En ocasiones las evaluaciones eran un mero requisito, y en otras formaban parte de una estrategia de comunicación social, pero no tuve la fortuna de ver que se usaran con fines de aprendizaje.

Por estos motivos, me interesó conocer cómo evaluaban su trabajo diferentes organizaciones con proyectos en una reserva de la biosfera. ¿Cómo se daban cuenta si sus proyectos beneficiaban a las comunidades? ¿Qué hacían con esa información? De esta manera, al enfocar mi interés en las organizaciones, y no en sus proyectos, la investigación se circunscribió en un área poco explorada en el estudio de evaluaciones en ANP: el uso de evaluaciones.

La evaluación de ANP en particular, y del manejo de recursos naturales en general, es un campo de generación de conocimiento muy amplio y dinámico. En él, es posible distinguir tres vertientes en función de su objeto de estudio (Cuadro 1).

Cuadro 1. Diferentes formas de abordar la evaluación de ANP.

OBJETIVO	OBJETO DE ESTUDIO
1. Producir evidencia para evaluar la efectividad de las ANP.	Políticas, programas y proyectos
2. Analizar las metodologías de evaluación en uso.	Metodologías
3. Estudiar el uso de evaluaciones.	Organizaciones

Fuente: Elaboración propia.

La primera vertiente genera evidencia sobre variables vinculables a la efectividad de las ANP. En ese sentido, las investigaciones evalúan el estatus o tendencias en indicadores de integridad ecológica como el cambio de uso de suelo, la dinámica de especies clave, o la reducción de presiones y amenazas (Andam *et al.*, 2008; Sánchez *et al.*, 2009; Ramírez *et al.*, 2011; Cuervo y Monroy, 2012; De la Rosa, 2016; Huitrón, 2019). A su vez, los estudios valoran los impactos socioeconómicos generados, como incrementos en los ingresos, cambios en el bienestar, o consecuencias no intencionadas, como la captura de beneficios por parte de grupos con mayor poder adquisitivo y capital social (Craig *et al.*, 2010; Vázquez y Fermán, 2010; Mckinnon *et al.*, 2016; Friedman *et al.*, 2018).

El segundo enfoque, orientado hacia objetivos metodológicos, busca mejorar la forma en la que se genera evidencia creíble sobre la obtención de resultados ambientales. Por ello, las investigaciones promueven debates epistemológicos y éticos sobre las metodologías en uso⁶ (Roetman, 2011; Richards, 2012; Roe *et al.*, 2013; Baylis *et al.*, 2016), perfeccionan los indicadores utilizados (Woodhouse *et al.*, 2015; Game *et al.*, 2018, Woodhouse *et al.*, 2018), clarifican conceptos básicos (Salafsky *et al.* 2008; Margoluis *et al.*, 2013; Pressey *et al.*, 2015; Salafsky *et al.*, 2019), o desarrollan nuevas metodologías de evaluación (Ervin, 2003; Pomeroy *et al.*, 2004; Hockings *et al.*, 2006; Stolton y Dudley, 2016).

Finalmente, el tercer enfoque estudia los procesos sociales que influyen en la producción y uso de evaluaciones. Después de todo, si es tan importante tomar decisiones basadas en evidencia, entonces se necesita explorar si está ocurriendo o no (Wild y Ramalingam, 2018). Más allá de análisis metodológicos sobre su calidad, este tipo de investigación explora por qué las organizaciones realizan cierto tipo de evaluaciones, si las usan y para qué las usan. La relevancia del uso, frecuentemente olvidada, se encuentra en que el valor de una evaluación no está en su producción *per se*, sino en su uso para la implementación de iniciativas que sean cada vez más efectivas (Feinstein, 2001; Stoll, 2010). En esta vertiente se inserta la tesis doctoral que se presenta aquí.

En este tipo de investigación se yuxtaponen disciplinas centradas en el estudio de las organizaciones con disciplinas que se enfocan en la interfaz sociedad-naturaleza, como la geografía ambiental. La identidad expansiva de esta última disciplina se considera una virtud, pues permite ser un ámbito no confinado a la propia geografía, contando con fronteras flexibles que promueven contribuciones divergentes para proponer respuestas más robustas en el estudio de la relación sociedad-naturaleza (Castree *et al.*, 2009).

Sin embargo, Castree, Demeritt y Liverman (2009) señalan que es preciso cuidar que la geografía ambiental no se convierta en un recipiente donde recaiga todo lo que puede ser dicho, recordando el estilo monográfico de escritura donde efectivamente se encuentra reunida toda la información sobre un tema, pero parcelada en secciones que no tienen posibilidades de diálogo entre sí. Por lo

⁶ En los extremos opuestos se encuentran metodologías rigurosas (basadas en diseños experimentales y cuasi-experimentales) que ponen énfasis en los resultados, y metodologías participativas que ponen énfasis en el proceso (Roetman, 2011).

contrario, los autores invitan a pensar la conexión de forma asimétrica y vertical, haciendo evidentes los puntos de contacto entre múltiples proyectos de investigación que consideran la relación entre sociedad y naturaleza, incluso fuera de los límites de la propia disciplina (Castree *et al.*, 2009: 6).

Este carácter poroso es palpable en áreas de investigación que conectan el ambiente con el desarrollo (Perreault, 2009), y que por ende estudian simultáneamente a) la relación de iniciativas de conservación con la obtención de resultados ambientales y b) el vínculo de procesos de desarrollo con resultados socio-económicos. Como campo interdisciplinario, su estudio abreva de las teorías y métodos provenientes de la geografía, sociología, antropología, economía, ciencias políticas y ciencias naturales para responder preguntas pragmáticas asociadas al cómo diseñar e implementar políticas exitosas (Bennett *et al.*, 2017). Por lo anterior, el manejo de ANP es uno de sus objetos de estudio paradigmáticos (Perreault, 2009).

La alarmante evidencia sobre el bajo impacto de las evaluaciones ha demostrado que, si bien son necesarias, no son una condición suficiente para que las organizaciones sean efectivas (Leverington *et al.*, 2010; Carranza *et al.*, 2014). Con frecuencia las evaluaciones son irrelevantes para tomar decisiones o afectan negativamente el comportamiento organizacional (Ebrahim, 2005; Guijt, 2008). La razón es que, lejos de ser solo un procedimiento técnico para establecer relaciones de causalidad, las evaluaciones son un proceso social que se inserta en el diverso y conflictivo mundo de la toma de decisiones (Weiss, 1999). Por ello, si se quiere usar las evaluaciones para cumplir los objetivos que se esperan de ellas, es necesario entender el contexto en el que se insertan (Taut, 2007).

En general, mejorar el desempeño a partir de evaluaciones es un ideal con pocas posibilidades de implementación si no se consideran los requerimientos organizacionales para que ocurra. Las metodologías suelen hacer énfasis en la información, suponiendo que con ella todo caerá en su lugar, pero soslayan los procedimientos, habilidades y valores que se requieren para integrar el conocimiento en los procesos de toma de decisión (Guijt, 2008; Wild y Ramalingam, 2018). Hay muy poca información disponible sobre cómo las organizaciones están utilizando evidencia para aprender y adaptar sus proyectos (Hernández *et al.*, 2019).

También es importante considerar el marco institucional en el que operan las organizaciones, pues con frecuencia esta esfera tiene el poder de definir o alterar las evaluaciones que utiliza una organización. Lamentablemente, las evaluaciones formativas están lejos de ser la norma en el campo del desarrollo. A pesar de que en el papel se aspira al manejo adaptativo de los proyectos, muchos gobiernos y donantes siguen favoreciendo evaluaciones que miden la obtención de objetivos preestablecidos (Guijt, 2008).

En ese sentido, con frecuencia las evaluaciones se reducen a reportes de rendición de cuentas que no son útiles para inferir la generación de beneficios de conservación y desarrollo (Stem *et al.*, 2005; Kapos *et al.*, 2009). Además, ese tipo de evaluación es una barrera para el manejo adaptativo de los proyectos, dado que

las organizaciones temen las consecuencias en caso de reportar resultados distintos a los esperados, por ejemplo, perder la reputación sobre la capacidad de generar resultados, aspecto en el que se basa buena parte del éxito de la gestión de recursos (Hockings *et al.* 2006; Leverington *et al.*, 2010; Craigie *et al.*, 2015; McKinnon *et al.*, 2015). Dado este marco institucional, las organizaciones podrían tener incentivos para utilizar las evaluaciones como un medio de construcción de credibilidad y no necesariamente como un mecanismo para desempeñarse mejor (Mayne, 2006; Agudo, 2015).

Dicho lo anterior, el **objetivo de la investigación** fue comparar la forma en que distintas organizaciones hacen y usan evaluaciones en una reserva de la biosfera del centro de México. En específico, como casos de estudio se eligieron una administración municipal, una dependencia del gobierno federal, una universidad pública y una organización de la sociedad civil que tienen proyectos vinculados con el manejo de recursos naturales.

Precisamente por la posibilidad de generar consecuencias negativas como resultado de su involucramiento en el estudio, se tomó la decisión de garantizar el anonimato de las organizaciones. Por ende, tampoco se especifica la reserva de la biosfera en la que laboran. Esta elección, fundada en estándares éticos de investigación en ciencias sociales, busca respetar los intereses y privacidad de los involucrados en el estudio. En la metodología del estudio se amplían las razones de esta decisión.

Si bien se eligieron cuatro casos por su potencial de arrojar insumos relevantes para explorar el problema de investigación, más casos habrían brindado un espectro explicativo mayor. Sin embargo, hay consideraciones pragmáticas asociadas al tiempo y presupuesto que obligan a limitar el universo de estudio.

Cuando se abordan temas tan sensibles como las evaluaciones, hay que tomar en consideración que no es posible predecir el nivel de acceso que se tendrá. Como quedará claro en la descripción de la metodología, un análisis de los recursos disponibles y de la capacidad para acceder a la información en las organizaciones condujo a la selección de dos estudios de caso centrales y dos estudios de caso secundarios.

En los casos principales (dependencia federal y universidad pública) fue posible realizar observaciones a profundidad del fenómeno, y, en el caso de la universidad pública, interactuar con la finalidad de promover cambios en su sistema de evaluación. Mientras tanto, los casos secundarios (administración municipal y organización de la sociedad civil) brindaron elementos de contraste para comparar variables clave en distintas configuraciones organizacionales. Esta decisión da respuesta al riesgo de sacrificar (*trade-off*) descripciones detalladas cuando se hacen estudios comparativos en contextos donde, de cualquier modo, habría mérito en invertir todos los recursos en un solo caso.

En el análisis comparativo, se estudiaron factores organizacionales e institucionales con influencia en las características, supuestos y uso de las evaluaciones, poniendo especial atención en las barreras para que los integrantes de las organizaciones usaran evaluaciones para aprender. En ese sentido, la

investigación no evaluó la efectividad de sus proyectos, sino que buscó explorar las causas por las que las organizaciones utilizaban, o no, evaluaciones como una herramienta de aprendizaje.

La tesis atiende cuatro **brechas en la comprensión del fenómeno** que justifican su realización. Estas fueron identificadas al recuperar el conocimiento acumulado en torno del objeto de estudio.

1. A nivel internacional, se han hecho llamados para comprender mejor cómo se manifiesta en la práctica el manejo adaptativo en el campo del desarrollo. A su vez, se ha solicitado incrementar la base de evidencia sobre las condiciones necesarias para institucionalizarlo, de forma que logre sus objetivos de iteración programática y toma de decisiones basadas en evidencia. Por ello, se requieren más estudios de caso que ilustren el fenómeno (Guijt, 2008; Wild y Ramalingam, 2018; Pasanen *et al.*, 2019).

2. En México, no se ha estudiado la forma en que las universidades públicas evalúan sus proyectos de investigación aplicada⁷. El énfasis se ha puesto, en cambio, en el sistema de evaluación de la política científica federal y su influencia sobre el comportamiento de los investigadores (Galaz y Gil, 2013; Suárez y Muñoz, 2004).

3. No hay investigaciones sobre el sistema de evaluación diseñado por la Comisión Nacional de Áreas Naturales Protegidas (CONANP), dependencia federal responsable del manejo de ANP en México. Dado que normalmente se considera que los usuarios de las evaluaciones son funcionarios de gobierno de primer nivel (Cardozo, 2015), las investigaciones se han centrado en esa escala, al estudiar las evaluaciones externas de sus programas de subsidio (Orta *et al.*, 2013; Sotelo *et al.*, 2015).

4. En México, el estudio sobre el uso de evaluaciones se ha concentrado en la política de evaluación del gobierno federal (Cejudo y Maldonado, 2011; Cardozo, 2015). En su estudio, han sido pioneros tanto el Centro para el Aprendizaje en Evaluación y Resultados de América Latina y el Caribe (CLEAR-LAC) como el Centro de Investigación y Docencia Económicas (CIDE). Sin embargo, más allá de investigaciones que abordan los sistemas de evaluación de entidades federativas y de algunos municipios capitales (Maldonado y Galíndez, 2013; Consejo Nacional de Evaluación de la Política de Desarrollo Social, 2016) existen pocas investigaciones sobre el uso de evaluaciones en la escala local, con organizaciones directamente involucradas en la implementación de proyectos de conservación y desarrollo. Es importante estudiar el fenómeno a ese nivel, pues los integrantes que operan las intervenciones deberían ser los principales usuarios de evaluaciones para manejar a través de adaptaciones.

Específicamente, la tesis se basó en las siguientes preguntas de investigación:

⁷ Si bien, toda investigación busca generar un cambio al aportar nuevos conocimientos, se considera que una investigación es aplicada cuando existe una intencionalidad expresa de utilizar los conocimientos generados para atender problemas sociales.

Pregunta 1 (Capítulo 4): En la dependencia federal ¿Cómo se hacen y usan evaluaciones relacionadas con el manejo del ANP?

- a. ¿En qué grado las evaluaciones se han utilizado para el manejo adaptativo del ANP?
- b. ¿Cuáles son las principales barreras para que los guardaparques aprendan a partir de esas evaluaciones?

Pregunta 2 (Capítulo 5): En la universidad ¿Cómo se hacen y usan evaluaciones en un centro público de investigación agrícola inserto en la reserva?

- a. ¿Qué factores influyen en el propósito de las evaluaciones utilizadas por la universidad?
- b. ¿Cómo influye el sistema de evaluación sobre las decisiones de los investigadores?

Pregunta 3 (Capítulo 6): ¿Cómo se hacen y usan evaluaciones en cuatro organizaciones distintas con proyectos en la reserva?

- a. ¿Cuáles son sus semejanzas y diferencias?
- b. ¿Qué factores promueven o inhiben el uso de evaluaciones como una herramienta de aprendizaje organizacional?

La estructura del documento es la siguiente. El primer capítulo presenta el marco teórico de la investigación. En él, se describen elementos básicos del campo de la evaluación haciendo referencias constantes a los campos de la conservación y el desarrollo. Además, se presenta una taxonomía de las diferentes formas en que las evaluaciones pueden usarse y una síntesis de las variables relacionadas con el uso en la literatura sobre el tema.

El segundo capítulo desarrolla la metodología con la que se cumplió el objetivo de la investigación. En él, se exponen los distintos enfoques a través de los que se ha abordado el objeto de estudio y con base en ello se justifica la relevancia de los métodos seleccionados. En un segundo momento se describe cada uno de los componentes de la investigación y se presenta un análisis de las fortalezas y limitaciones del estudio.

El tercer capítulo describe el área de estudio, describiendo características de la reserva de la biosfera que permiten comprender su dinámica ambiental. Además, se caracteriza cada una de las organizaciones estudiadas y se perfilan elementos para comprender sus sistemas de evaluación.

Los resultados de la investigación se presentan divididos en tres artículos, dos de ellos publicados y uno en proceso de publicación. Por este motivo, 1) existe en ellos cierto grado de repetición del material presentado en el cuerpo de la tesis y 2) la numeración de apartados, cuadros, figuras y notas al pie siguen el orden interno de los artículos, sin fluir con la numeración del resto del texto de la tesis. Para facilitar al lector la identificación de estas diferencias, se tomó la decisión de distinguir la fuente empleada en esos capítulos⁸.

El cuarto capítulo presenta los resultados de la investigación que dan respuesta a la pregunta 1. De este modo, el artículo publicado en la revista *Sociedad y Ambiente* “*¿Sirven las evaluaciones para aprender? Influencia de las evaluaciones sobre las decisiones de manejo en un Área Natural Protegida*” describe varios factores que promueven e inhiben el uso de evaluaciones para dirigir la toma de decisiones de los guardaparques que administran la reserva.

El quinto capítulo presenta los resultados de la investigación que dan respuesta a la pregunta 2. El documento “*El uso de evaluaciones en un centro público de investigación agrícola*” analiza las evaluaciones con las que interactúa un Centro Público de Investigación Agrícola que opera dentro de la reserva. En ese sentido, se incluyen tanto las evaluaciones externas de sus proyectos financiados con recursos públicos, como el sistema interno de evaluación de la universidad a la que pertenece.

El sexto capítulo presenta los resultados de la investigación que dan respuesta a la pregunta 3. De este modo, el artículo publicado en la revista *Gestión y Ambiente* “*¿Medimos lo que es valioso o valoramos lo que es medible? Comparación interorganizacional del uso de evaluaciones en una reserva de la biósfera*” presenta un análisis comparativo de la forma en la que las cuatro organizaciones estudiadas hacen y usan evaluaciones.

Finalmente, en el séptimo capítulo se sintetizan las principales contribuciones de la tesis, entremezcladas con recomendaciones específicas y vetas de exploración futura.

⁸ En ellos se emplea la fuente Times New Roman.

Capítulo uno

Enmarcando el problema del uso de evaluaciones en el campo de la conservación.

Give me the fruitful error any time, full of seeds, bursting with its own corrections.

-Vilfredo Pareto-

But be aware of the fallacy of failure. It is celebrated only when you succeed.

-Bruce Nussbaum-

1.1 CONCEPTOS BÁSICOS DE EVALUACIÓN

En política, las evaluaciones sirven para establecer la relación causal entre los medios y los fines públicos esperados (Sanderson, 2000). Su objetivo es eminentemente práctico, pues la intención es utilizar ese conocimiento para emitir juicios y proponer mejoras en proyectos, programas y políticas⁹ (Scriven, 2007). En ese sentido, el acto de evaluar implica dar por sentado que el error es consustancial a la actividad humana, y por tanto, que los proyectos y las personas pueden fallar. También, implica asumir que es racional incorporar los errores, y su posibilidad, al tomar decisiones.

En la administración pública, este principio de falibilidad erosionó la validez del control jerárquico centralizado. Dada la complejidad del mundo, es evidente que seguir instrucciones no conduce necesariamente al cumplimiento de objetivos (González y Cerdán, 2008). Por ello, desde el auge de la Nueva Gestión Pública¹⁰ en la década de los setenta, las evaluaciones se consideran una práctica clave para fundamentar la gestión de proyectos sobre bases racionales. Hoy en día, se espera que las evaluaciones acompañen el funcionamiento de todas las organizaciones ligadas a la generación de valor público, incluidas las OSC y las universidades públicas.

A continuación, se desarrollan una serie de conceptos básicos de la práctica evaluativa. En primer lugar, se presentan las categorías e implicaciones metodológicas de la función elemental de las evaluaciones: el establecimiento de la causalidad. En segundo lugar, se distinguen los tipos de evaluación en función del momento en el que se realizan. Finalmente, se desarrollan los diferentes propósitos por los cuales las organizaciones pueden realizar evaluaciones.

1.1.1 El establecimiento de la causalidad

En esencia, las evaluaciones buscan establecer si un proyecto consiguió los objetivos para los que fue diseñado. Por ello, su función es identificar los cambios generados y probar que estos son atribuibles a las acciones realizadas (Chelimsky, 1997). En su mínima expresión, las relaciones de causalidad de los proyectos se pueden representar como proposiciones condicionales ($p \rightarrow q$) donde las acciones son el antecedente (p) y los resultados el consecuente (q). La función de las

⁹ Cualquiera de estas categorías puede ser el objeto de las evaluaciones. Los proyectos, unidad mínima de asignación de recursos, pueden anidarse en programas, las cuales a su vez pueden ser parte de políticas públicas diseñadas para solucionar problemas específicos. A partir de ahora, para simplificar la redacción, esas tres categorías se sintetizarán en el concepto proyecto.

¹⁰ Los promotores de la Nueva Gestión Pública (New Public Management) criticaron el modelo burocrático de organización y operación estatal, orientado hacia la valoración del procedimiento y no a la producción de resultados de valor (González, 2013). Desde su auge, no solo se requiere revisar que los fondos públicos se utilicen siguiendo ciertas reglas, también hay que garantizar que se usa de forma eficaz y eficiente (Eyben, 2015).

evaluaciones es identificar el valor de verdad (verdadero/falso) de las proposiciones (p), (q) y (p→q) (Cuadro 2).

Cuadro 2. Tres escenarios donde la evaluación prueba las proposiciones de un proyecto.

$p \rightarrow q$	Proposición condicional que representa la hipótesis de un proyecto de desarrollo. La lectura de dicha proposición sería: Si (p) el proyecto implementa correctamente las acciones previstas, entonces (q) se obtendrán los resultados esperados.
Escenario 1. La evaluación prueba que...	
$\sim p$	No se realizaron todas las actividades necesarias o no se ejecutaron correctamente
$\sim q$	No se obtuvieron los resultados esperados
Escenario 2. La evaluación prueba que...	
p	Se realizaron correctamente todas las actividades previstas
$\sim q$	No se obtuvieron los resultados esperados
\therefore	Por lo tanto
$\sim(p \rightarrow q)$	Es falso que si se realizan correctamente todas las actividades entonces necesariamente se obtienen los resultados esperados (la hipótesis del proyecto se refuta)
Escenario 3. La evaluación prueba que...	
p	Se realizaron correctamente todas las actividades previstas
q	Se obtuvieron los resultados esperados
\therefore	Por lo tanto
$p \rightarrow q$	Es verdad que si se realizan correctamente todas las actividades entonces se obtienen los resultados esperados (la hipótesis del proyecto se valida)

Fuente: Elaboración propia

La tabla representa un patrón de causalidad simple, donde una causa se asocia a un resultado. Sin embargo, en proyectos de desarrollo esto puede ser mucho más

complejo, pues una o varias causas se pueden asociar a múltiples resultados (Stern *et al.*, 2012). A lo largo de los años se ha desarrollado un aparato conceptual para clarificar el tipo de relaciones de dependencia que pueden existir y probarse en los proyectos. A continuación, se presenta una tabla que sintetiza los principales conceptos (Cuadro 3) y, después, se incluye un ejemplo proveniente del campo de la conservación que esquematiza las relaciones de causalidad en proyectos de esta índole (Figura 1).

Cuadro 3. Terminología básica útil para establecer relaciones de causalidad.

TÉRMINO	DEFINICIÓN
Insumos (<i>Inputs</i>)	Son las inversiones necesarias para ejecutar las acciones. Estas pueden referirse tanto a los materiales necesarios como al tiempo del personal. Si bien, idealmente los insumos se determinan después de establecer los resultados que se pretende generar, normalmente ocurre al revés. Es decir, existe un presupuesto preestablecido que suele limitar las acciones potenciales que se pueden seleccionar. En ese sentido, los insumos tienen una influencia importante sobre los impactos alcanzables.
Acciones	Son las actividades concretas que hará un proyecto para atender un problema y conseguir un resultado. En el lenguaje científico, se consideran los tratamientos o variables independientes cuyo efecto se pretende demostrar.
Estrategias	Agrupan una serie de acciones afines para conseguir un resultado. Por ejemplo, una estrategia de comunicación puede implicar tres acciones: la realización de talleres, la difusión de información en redes sociales y la publicación de artículos de divulgación.
Productos o Resultados inmediatos (<i>Outputs</i>)	Son los productos concretos y contables que indican la realización exitosa de las acciones. Debido a que pueden estimarse con facilidad y certidumbre, se utilizan como un mecanismo para monitorear la implementación de los planes acordados. Algunos ejemplos pueden ser: evidencias de la construcción de cisternas de ferrocemento, una relación de personas capacitadas y un reporte que contenga la información solicitada. A pesar de su popularidad como métrica de evaluación, estos no informan si el proyecto generó el valor público esperado.
Efectos o Resultados de mediano plazo (<i>Outcomes</i>)	Se refieren a los cambios que se espera sean ocasionados gracias a la obtención de los productos de una intervención. Estas modificaciones pueden ser cambios en las habilidades y decisiones de actores clave o bien cambios en variables ambientales de interés. Por ejemplo: <ul style="list-style-type: none"> • Gracias a una serie de capacitaciones (<i>output</i>) el 29% de los ganaderos de la región han adoptado técnicas de pastoreo racional (<i>outcome</i>). • Gracias a las obras de conservación de suelos en la unidad de escurrimiento (<i>output</i>) ha disminuido un 19% el volumen de sedimentos que se vierten a la presa (<i>outcome</i>).

	<p>En el lenguaje científico, estos cambios se conciben como las variables dependientes que son afectadas por las acciones. La medición de estos cambios puede ser suficiente para estimar el valor público generado, aunque hay que tener claro que no es así en todos los casos. Por ejemplo, un proyecto puede lograr incrementos en el ingreso de poblaciones vulnerables, pero eso resultará en un impacto social positivo en función de la forma en que se distribuya y gaste ese ingreso.</p>
Impactos o Resultados de largo plazo	<p>Son las postulaciones aspiracionales de los logros a conseguir. Por ello, suelen ser declaraciones cualitativas de la consecuencia intencionada. Por ejemplo:</p> <ul style="list-style-type: none"> • En la cuenca del Pánuco se han construido los acuerdos sociales necesarios para manejar el agua de forma sostenible. En ese sentido, las fronteras políticas (municipios y estados) e hidrológicas (microcuencas, subcuencas, acuíferos) se coordinan para tomar decisiones en la escala administrativa relevante.
Resultados y metas	<p>Es importante tener claro que en el campo del desarrollo existe ambigüedad en el uso de estos vocablos. De manera general, son términos genéricos que se usan para describir un estado futuro deseado. La confusión proviene de que tanto impactos como <i>outcomes</i> y <i>outputs</i> pueden considerarse resultados o metas. En esta tesis, se usa la palabra resultados sólo para hacer referencia a la generación de valor público resultante de un proyecto, es decir un <i>outcome</i> o un impacto.</p>
Indicadores	<p>Son unidades de medida monetizables, cuantitativas, o cualitativas que permiten a las evaluaciones identificar avances en el cumplimiento de un <i>output</i> o de un <i>outcome</i>. Por ejemplo:</p> <ul style="list-style-type: none"> • # de capacitaciones brindadas a productores (<i>output</i>) • % de productores que incorporan la tecnología transferida en su sistema productivo (<i>outcome</i>) • % de incremento en las ganancias obtenidas por los productores (<i>outcome</i>)
Supuestos	<p>Pueden referirse a las hipótesis sobre las relaciones de causalidad ($p \rightarrow q$) de un proyecto, las cuales se hacen explícitas en diagramas que representan la conexión entre acciones, <i>outputs</i>, <i>outcomes</i> e impactos.</p> <p>A su vez, pueden referirse a supuestos sobre el estatus de variables externas que no son afectadas por la intervención, pero de las cuales depende su éxito. Por ejemplo, un proyecto de reforestación con especies nativas puede diseñarse asumiendo que a los propietarios de los terrenos les interesan las especies provistas por el proyecto y que les interesa tener una plantación forestal.</p>
Cadena de resultados	<p>Son diagramas que representa la secuencia lógica mediante la que se espera que un cambio ocurra.</p>
Partes interesadas (stakeholders)	<p>Son una organización, grupo o individuo que tiene un interés directo o indirecto sobre una iniciativa de desarrollo, o quién afecta o es afectado positiva o negativamente por un resultado de ella.</p>

Relevancia	El grado en el que los objetivos de un proyecto son consistentes con las prioridades del grupo objetivo.
Sostenibilidad o Robustez	La probabilidad con la que los efectos positivos de la iniciativa persistirán por un periodo extendido después de que las actividades del proyecto terminen.

Fuente: Elaborado a partir de Kusters *et al.* (2011), Margoluis *et al.* (2013) y Pressey *et al.* (2015).

Figura 1. Cadena de resultados de un proyecto de conservación



Fuente: Modificado de Margoluis *et al.* (2013).

Dado que el objetivo básico de las evaluaciones es aportar información veraz sobre las relaciones de causalidad, es valiosa la existencia de un debate metodológico orientado a la mejora de este propósito. Es posible contrastar enfoques con diferentes raíces éticas y epistemológicas (Roetman, 2011), y que, por tanto, tienen diferentes visiones sobre qué se ha de evaluar, cómo se hará y para qué se utilizará el conocimiento generado (Shadish, 1998 en Donaldson y Assam, 2015).

Uno de los fundamentos que distinguen entre sí a los enfoques evaluativos es el tipo de inferencia causal que utilizan. En las ciencias sociales existen cuatro maneras de mirar la causalidad (Befani, 2012). En primer lugar, se encuentra la inferencia por regularidad, la cual se basa en frecuencias de asociación entre causas y efectos. Esta es la base de enfoques que se apoyan en análisis estadísticos. En segundo lugar, se encuentra la inferencia contrafactual, la cual usa un grupo de control para aislar el efecto de una intervención. En ella, se hace una comparación entre dos casos idénticos excepto porque uno recibió un tratamiento y el otro no. Si la comparación da como resultado diferencias significativas entre los dos grupos, las diferencias pueden ser atribuibles a la intervención y no a otras influencias. Esta es la base de los enfoques que emplean diseños experimentales y cuasi-experimentales.

En tercer lugar, se encuentra la causalidad múltiple, en donde la inferencia se basa en la valoración de combinaciones de causas que explican el efecto. Esta es la base de enfoques configuracionales de evaluación de impacto como el análisis cualitativo comparativo (QCA por sus siglas en inglés). Finalmente, se encuentra la causalidad generativa, en donde se basa la inferencia en la identificación de los mecanismos¹¹ que explican los efectos. Este tipo de inferencia es utilizada por enfoques de evaluación basados en teorías (Westhorp, 2014).

Entre los tipos de inferencia mencionados, la contrafactual es muy valorada para evaluar políticas públicas. Esto es así pues se percibe que las evaluaciones basadas en diseños experimentales cumplen cabalmente con todos los principios de integridad metodológica, lo que garantiza su objetividad y rigor (Campbell y Stanley, 1978) (Cuadro 4). Por ello, con frecuencia se les considera el estándar dorado para estimar los impactos generados (Torres y Preskill 2001; Richards, 2012; Visser *et al.*, 2013).

¹¹ Un mecanismo se compone de entidades (por ejemplo, personas, organizaciones o sistemas) que llevan a cabo actividades (por ejemplo, protestar, investigar, proyectar) para generar un resultado. Su funcionamiento está ligado a un contexto, siendo por tanto dependientes de una serie de condiciones de arranque y de la existencia de factores de soporte (Stern *et al.*, 2012; Centre for Development Impact, 2015).

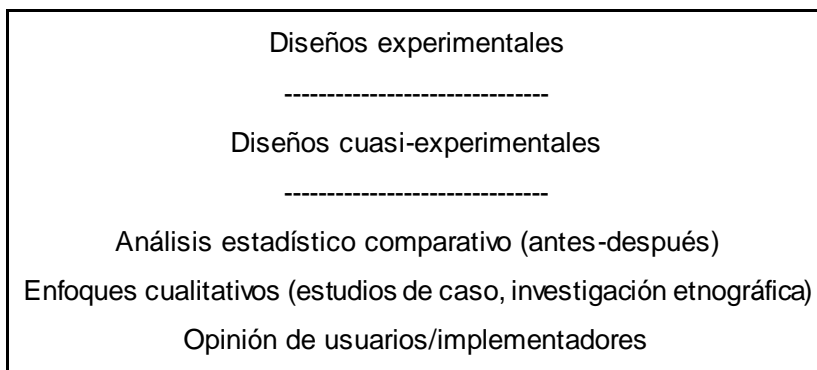
Cuadro 4. Principios para la integridad metodológica de evaluaciones de impacto.

PRINCIPIO	DESCRIPCIÓN
Credibilidad	Desde la perspectiva de su audiencia las conclusiones de la evaluación son creíbles y legítimas.
Validez externa	Los resultados pueden extenderse a grupos y contextos distintos de los que son estudiados.
Validez interna	Los resultados establecen una relación causal entre el programa y sus efectos.
Validez de la medición	La medición de la relación causal es precisa debido a que tanto las técnicas de medición, la frecuencia de medición y los indicadores seleccionados son adecuados.
Confiabilidad	Los procedimientos de medición producen resultados similares en observaciones repetidas de la misma condición o evento.

Fuente: Hatry y Newcomer (2010)

Debajo de ella se escalonan otras metodologías por su capacidad para producir conocimiento seguro sobre las relaciones de causalidad en las intervenciones (Cuadro 5). En consecuencia, organismos multilaterales y dependencias federales encargadas de la evaluación de programas sociales privilegian el uso de esa metodología, considerándola con frecuencia, como la única válida (CONEVAL, 2007; Gertler *et al.*, 2017).

Cuadro 5. Jerarquía de fuentes de información de conocimiento seguro sobre las relaciones de causa-efecto en las intervenciones.



Fuente: Modificado de Vedung, 2010

En este orden de ideas, en los últimos años se han hecho llamados para incrementar ese tipo de evaluaciones en el campo de la conservación (Pressey *et al.*, 2015; Baylis *et al.*, 2016). El cuadro 6 ejemplifica cómo distinguir entre *outcomes* e impactos en proyectos de conservación que utilizan la inferencia contrafactual.

Cuadro 6. Diferencia entre outcomes e impactos en la inferencia contrafactual.

	Outcome	Impacto
tiempo ₀	condición _{p0}	condición _{p0} – condición _{u0}
tiempo ₁	condición _{p1}	condición _{p1} – condición _{u1}
tiempo ₂	condición _{p2}	condición _{p2} – condición _{u2}
tiempo ₃	condición _{p3}	condición _{p3} – condición _{u3}
⋮	⋮	
tiempo _n	condición _{pn}	condición _{pn} – condición _{un}
<p>Un <i>outcome</i> se obtiene al generar cambios en las condiciones de un objeto de conservación (p), los cuales se registran en múltiples momentos, revelando tendencias (p. ej. pasar de la condición_{p0} a la condición_{p2}). En concreto, esto puede ser la disminución de una amenaza que presionaba la población de la especie o mejoras en su hábitat.</p>		
<p>Un impacto es la diferencia entre las condiciones conseguidas en el área de influencia del proyecto (p) y las condiciones de sitios en donde el proyecto no tiene influencia (u), es decir el escenario contrafactual.</p>		

Fuente: Modificado de Pressey *et al.* (2015)

Si bien se considera como el método más riguroso para establecer la causalidad, esto no implica que esté libre de limitaciones. Los diseños (cuasi-) experimentales han sido criticados por su alto costo y por problemas éticos y epistemológicos asociados con el aislamiento de controles (Richards, 2012; Befani *et al.*, 2014).

Por otro lado, se ha señalado que son incapaces de explicar cómo se generaron los efectos (Sanderson, 2000; Visser *et al.*, 2013; Van Hemelrijck y Guijt, 2016). Esto es así, pues si bien la inferencia contrafactual establece la relación lineal entre una intervención y una serie de cambios, no es capaz de identificar los mecanismos causales subyacentes que explican por qué ocurrieron. En este sentido, la implementación de los proyectos es una caja negra, y la influencia del contexto

organizacional y político pasa desapercibida. En general, esto merma la capacidad de los diseños experimentales para identificar áreas específicas de mejora en los proyectos (Mackay y Horton, 2003).

Gran parte de los proyectos de desarrollo operan en circunstancias complejas, donde, dada su incertidumbre, es importante entender por qué, o por qué no, funcionan los proyectos en escenarios específicos (Visser *et al.*, 2013). Esto implica entender que el contexto influye tanto en lo que un mecanismo “es” como sobre su funcionamiento (Cuadro 7).

Cuadro 7. Influencia del contexto sobre la efectividad del mecanismo.

Contexto simple	Mecanismo → <i>Outcome</i>	El resultado no depende del contexto. Ejemplo: La efectividad de una vacuna para proteger contra enfermedades no está en función del lugar donde se aplica.
Contexto complejo	Contexto + Mecanismo → <i>Outcome</i>	El resultado depende de la interacción del mecanismo con el contexto. Ejemplo: La efectividad de un programa de reforestación en zonas indígenas está en función de que se adapte a las condiciones locales.

Fuente: Elaboración propia a partir de Olsen, 2014

Por ello, con frecuencia se recurre a metodologías cualitativas que, aunque sean menos rigurosas para establecer la causalidad y tengan una baja validez externa, son adecuadas para responder las preguntas anteriores al abrir la “caja negra” de la implementación de los proyectos (Sanderson, 2000; Perrin, 2002).

Estas metodologías se basan en teorías que esquematizan las cadenas plausibles de resultados, es decir, predicen los vínculos entre los productos, los efectos directos y los impactos (Richards, 2012). En este contexto, la función de las evaluaciones es identificar los cambios y elucidar los factores que los explican, probando la verdad de las teorías. Los modelos de representación más conocidos en el campo del desarrollo son el marco lógico (Levine, 2007; Department for International Development, 2011) y la teoría del cambio (Funnell y Rogers, 2011).

Para establecer los cambios, estos enfoques necesitan líneas de base para comparar el antes y el después del proyecto, pudiendo utilizar una mezcla de datos cuantitativos y cualitativos. Además, con base en un escenario social de referencia hipotético, contrastan los resultados con cómo hubiera cambiado la situación en un escenario “sin proyecto” (Richards 2012; Bell y Aggleton, 2012). En metodologías participativas, este proceso es liderado por los mismos beneficiarios, siendo ellos mismos quienes valoran los cambios.

Una de las debilidades de los enfoques de evaluación basados en teorías es que puede esquematizarse de forma muy simple la relación entre causas y efectos.

Además, las teorías pueden ser sujetas a presión política, y en lugar de considerarse como una hipótesis a ser probada, ser la historia oficial sobre los cambios conseguidos. En esos casos, las evaluaciones son meros simulacros que corroboran las teorías, y no investigaciones para establecer lo que realmente ocurrió con las personas (Batiwala y Pittman, 2010 en Bell y Aggleton, 2012).

En última instancia, la selección del método está en función de la base de recursos disponible, la escala del objeto de evaluación (política, programa o proyecto) y el tipo de evidencia requerido. Ningún método es mejor en sí, cada enfoque es valioso cuando se utiliza en circunstancias que se adecúan a sus fortalezas e intenciones (Chelimsky, 1997; Salafsky *et al.*, 2019). De hecho, se ha resaltado la complementariedad de las metodologías para robustecer las evaluaciones de impacto. Los datos cualitativos permiten identificar las condiciones bajo las cuales los impactos surgen. Los datos cuantitativos, por su parte, permiten establecer muestras representativas sobre las cuales estimar la magnitud de los efectos (Baylis *et al.*, 2016).

Esta visión holística implica trascender los supuestos que consideran a los hallazgos cuantitativos como la única evidencia empírica válida y a los hallazgos cualitativos como juicios de valor que poco aportan a la comprensión del mérito de un proyecto (Stern *et al.*, 2012; Agudo, 2015). Después de todo, tanto las evaluaciones cualitativas como las cuantitativas son susceptibles a múltiples sesgos cognitivos y a presiones políticas, sociales y económicas (Camfield *et al.*, 2014).

El rigor no lo otorga el uso de un determinado método, en cambio, este se logra garantizando la calidad del pensamiento puesto en el diseño y en la implementación de una metodología. Esto implica que un evaluador debe:

- Ser reflexivo sobre los supuestos que están detrás de la metodología de evaluación elegida (Guijt, 2008),
- Ser transparente respecto de su posición e interés (Camfield *et al.*, 2014),
- Ser consciente de los intereses que están alrededor de la evaluación (Weiss, 1999),
- Reconocer la probabilidad del sesgo (Camfield *et al.*, 2014),
- Asegurar que el muestreo, la triangulación de métodos, la facilitación de procesos, la compilación de información, la validación cruzada y los análisis de causalidad se implementan de forma adecuada (Van Hemelrijck y Guijt, 2016).

Todo esto permite asegurar consistencia y responder a los propósitos y limitaciones de la evaluación, condiciones necesarias para que exista confianza sobre sus hallazgos (Van Hemelrijck y Guijt, 2016).

1.1.2 Clasificación de las evaluaciones por el momento en el que se llevan a cabo

En función del momento del ciclo de manejo de los proyectos en el que son efectuadas, las evaluaciones pueden clasificarse como evaluaciones *ex ante*, *ex post*, o simultáneas.

1.1.2.1. Evaluaciones *ex ante*

Cuando se realizan antes de emprender un proyecto, el propósito de las evaluaciones es analizar las consecuencias más probables de distintas alternativas de acción, con la finalidad de guiar la decisión. En el campo del desarrollo, estas evaluaciones, conocidas como *ex ante* o prospectivas, tienen el objetivo de valorar el impacto potencial de las intervenciones, sea como un criterio de elegibilidad¹² o para mejorar su diseño, de forma que se maximicen las posibilidades de generar los resultados esperados y se minimice el riesgo de causar impactos negativos (Vanclay, 2003; Esteves *et al.*, 2012).

En la legislación ambiental de distintos países, este tipo de evaluación se utiliza frecuentemente como una herramienta regulatoria. Su objetivo es cancelar proyectos que acarrearían daños significativos, o bien, garantizar que se toman medidas para prevenir o mitigar sus consecuencias negativas. En México, la Ley General del Equilibrio Ecológico y la Protección al Ambiente (LGEEPA) especifica el procedimiento para realizar evaluaciones de impacto ambiental a distintas obras o actividades, dentro de los cuales están cambios del uso de suelo de áreas forestales y acciones dentro de ANP federales (SEMARNAT, s.f.). Por su parte, tanto la Ley de Hidrocarburos como la Ley de la Industria Eléctrica, solicitan evaluaciones de impacto social para identificar externalidades sociales en proyectos de energía, y con base en ellas, generar planes de gestión para dar seguimiento a las medidas acordadas (Rousseau, 2017).

Un aspecto central en estas evaluaciones es tener un buen conocimiento de los intereses y necesidades de los diferentes grupos que serán afectados por los proyectos. Por ello, las metodologías exigen la participación de los actores locales en el proceso de identificación de los impactos, así como en el diseño de medidas para evitar, mitigar o compensar los posibles efectos negativos que proyectos de desarrollo pudieran ocasionarles (Richards y Panfil, 2011; Esteves *et al.*, 2012). Estas medidas precisan asegurar que grupos vulnerables (como los pueblos indígenas) reciben beneficios sociales y económicos apropiados culturalmente, siendo además incluyentes desde el punto de vista de género e intergeneracional (Banco Mundial, 2005). En la legislación mexicana, es obligatorio hacer consultas públicas en torno de las evaluaciones de impacto ambiental y, además, conducir un proceso de consulta previa, libre, informada cuando los proyectos se desarrollarán en los alrededores de comunidades indígenas.

¹² P. ej. para decidir qué proyectos financiar en función de su relación costo/beneficio.

Cada vez más, este procedimiento forma parte de los marcos normativos que regulan el flujo del financiamiento internacional de proyectos de desarrollo. En torno de proyectos de manejo de recursos naturales, las evaluaciones *ex ante* son un estándar en intervenciones REDD+ (Richards y Panfil, 2011; Mwampamba *et al.*, 2015) y para acceder al mercado voluntario de carbono (CCBA, 2013).

1.1.2.2. *Evaluaciones simultáneas, intermedias o de proceso*

Este tipo de evaluación se lleva a cabo durante la ejecución de un proyecto. Por un lado, son un mecanismo para reportar avances en la entrega de los *outputs* esperados. Por el otro, sirven para revisar si el desarrollo de las acciones tiene la eficiencia operacional esperada, información que se utiliza para tomar decisiones en tiempo real que se espera incrementen la efectividad del proyecto.

Por el segundo de sus propósitos, este tipo de evaluación es la base del manejo adaptativo de los proyectos. Esto es así, pues su objetivo es entender y responder mejor a los cambios que ocurren en el contexto de operación. Para lograrlo, las evaluaciones recurren tanto a métodos de recolección de información, como a métodos de aprendizaje organizacional para que el equipo haga sentido de lo que está sucediendo y registre las razones por las que se tomaron decisiones correctivas (Guijt, 2008; Patton, 2010).

Por sus características, existe una porosa frontera conceptual entre este tipo de evaluación, el monitoreo y la gestión del conocimiento (Guijt, 2008; Chaplowe y Cousins, 2016). Las evaluaciones suelen relacionarse con procesos de análisis y el monitoreo con ejercicios periódicos de recolección de información. Sin embargo, cuando las evaluaciones se realizan de forma poco frecuente, o inclusive sólo al terminar los proyectos, es el monitoreo y el uso de técnicas para hacer sentido de lo que está pasando lo que permite hacer correcciones sobre la marcha. Por ello, para Guijt:

“no se requiere una división estricta entre los dos términos en sus funciones (...). En lugar de referirnos a los términos de monitoreo, evaluación, o M&E, podríamos hablar en términos del sistema de aprendizaje y mejora, el proceso de rendición de cuentas financiera, o el proceso de reflexión estratégica. Así, mayor atención podría darse a definir qué información se busca y a través de ello desarrollar procesos que contribuyan al aprendizaje, en lugar de perpetuar distinciones poco claras.” (2008: 113).

Hoy en día, en el campo del desarrollo hay múltiples metodologías de evaluación ligadas al manejo adaptativo (Earl *et al.*, 2001; Patton, 2010) y varias específicas para el campo de la conservación (Pomeroy *et al.*, 2004; Hockings *et al.*, 2006; Conservation Measures Partnership, 2007; Stolton y Dudley, 2016;). Dada su orientación hacia el aprendizaje, incorporan procedimientos para entender qué está sucediendo y facilitar la toma de decisiones del equipo interventor.

1.1.2.3 Evaluaciones Ex post

Mientras que las evaluaciones *ex ante* miran al futuro y las evaluaciones simultáneas miran al presente, las evaluaciones *ex post* se enfocan en el pasado. Este tipo de evaluación valora si un proyecto generó los resultados previstos. Si se limita a la medición de los *output* comprometidos por un proyecto, la evaluación será del desempeño que tuvo el equipo interventor. En cambio, si su propósito es identificar la vinculación causal entre actividades, productos y efectos, explicando cómo funcionó la intervención y para quién (Barnett y Munslow, 2014), entonces es una evaluación de impacto.

1.1.3 Propósito de las evaluaciones

Los términos “evaluación sumativa” y “evaluación formativa” se acuñaron desde hace décadas para distinguir los dos propósitos centrales que puede tener una evaluación (Scriven, 1967). La primera, relacionada con el concepto de rendición de cuentas, implica hacer una evaluación para juzgar el mérito de un proyecto en términos de su desempeño, efectividad o impacto (Mackay y Horton, 2003). Se suele realizar de manera *ex post* para orientar la toma de decisiones de la instancia a la que se le rinden cuentas (usualmente el proveedor de los recursos económicos o un área superior en la jerarquía organizacional).

Mientras tanto, la segunda, relacionada con el concepto de aprendizaje, se realiza para obtener información que permita mejorar el desempeño de un proyecto. Puede hacerse de manera simultánea o *ex post*, y su usuario suele ser el mismo personal que opera el proyecto (Patton y Horton, 2009). A continuación, se desarrollan las características de cada una de ellas.

1.1.3.1 Evaluaciones sumativas

En general, se considera erróneo que los responsables de gobernar no expliquen públicamente su conducta. Dada la frecuencia con la que, por desgracia, eso ha ocurrido, existe una preocupación legítima por transparentar las acciones que buscan generar valor público (Bovens *et al.*, 2008). Eso ha situado a la rendición de cuentas como un principio fundamental tanto en la administración pública como en el campo del desarrollo (Alkin y Christie, 2004; Cousins *et al.*, 2015).

El significado dominante del concepto implica que:

“Cuando promesas se hacen o se reúnen recursos en nombre de un grupo de personas o de una causa, se considera justificado esperar una retroalimentación sobre cómo se usó el dinero y cómo se cumplieron las promesas (...) Si esa retroalimentación no sucede entonces podrían sobrevenir sanciones.” (Guijt, 2010: 283)

En ese sentido, la rendición de cuentas es un concepto que establece relaciones entre actores desiguales. Uno de ellos con el poder para imponer un sistema de

control y cumplimiento (o al menos con el derecho de exigir resultados), y otro con el deber de explicar sus acciones (Guijt, 2010; Cousins *et al.*, 2015). Al estructurar de esa forma la dinámica de las relaciones, la rendición de cuentas puede analizarse desde el modelo del principal-agente.

“El modelo del principal-agente representa la toma de decisiones en una relación, en la que una parte, el principal, considera entrar en un acuerdo contractual con otro, el agente, con la expectativa de que el agente elegirá subsecuentemente acciones que produzcan los efectos deseados por el principal (...) El problema de decisión del principal va más allá de encontrar a una persona cualificada, pues no tiene garantías de que el agente, una vez contratado, de hecho elegirá perseguir los intereses del principal, o que lo hará de forma eficiente.” (Moe, 1984 en Meier y Krause 2003: 7).

En este tipo de relación, el principal, que desea delegar autoridad para que se hagan actividades en su representación, se enfrenta al problema de que no tiene información completa sobre las acciones del agente. Por ello, con la finalidad de asegurar su cumplimiento, establece mecanismos para incidir sobre su comportamiento (Meier y Krause 2003: 8). El cuadro 8, pone algunos ejemplos de relación principal-agente en el campo del desarrollo.

Históricamente, las evaluaciones sumativas han sido uno de los principales mecanismos para estimular el cumplimiento de estándares de desempeño (Cousins *et al.*, 2015). En ese sentido, podría decirse que el interés por la rendición de cuentas ha creado la necesidad de hacer evaluaciones, incluso al nivel de que se le considera una de las raíces de la práctica evaluativa (Alkin y Christie, 2004).

Existen diferentes tipos de evaluaciones sumativas que puede utilizar el principal para llamar a cuentas a los agentes. En el nivel más básico y generalizado se realizan evaluaciones de desempeño para demostrar que el proyecto se condujo conforme a los planes acordados y que consiguió objetivos preestablecidos (Chelimsky, 1997; Ebrahim, 2005). Para ello, este tipo de evaluación se basa tanto en auditorías financieras que verifican el uso previsto de los fondos como en la solicitud parcial y final de pruebas de la generación de los productos esperados (*outputs*). La información para realizar este tipo de evaluaciones suele ser reportada por los mismos agentes evaluados y cotejada por evaluadores externos.



Cuadro 8. Ejemplos de relación principal-agente en el campo del desarrollo.

PRINCIPAL	AGENTE
Relación interorganizacional	
<i>Donante</i> <ul style="list-style-type: none"> • Agencias de cooperación internacional • Bancos de desarrollo • Instituciones de Asistencia Privada 	<i>Receptor/Instancia Ejecutora</i> <ul style="list-style-type: none"> • Dependencia de gobierno • OSC • Universidad Pública
Relación intraorganizacional	
<i>Área superior o responsable de evaluación en la estructura organizacional</i> <ul style="list-style-type: none"> • Direcciones de evaluación • Contraloría Municipal 	<i>Área subordinada en la estructura organizacional</i> <ul style="list-style-type: none"> • Direcciones de área, departamentos, puestos.
Relación intergubernamental	
<i>Instancia responsable de la evaluación del desempeño gubernamental</i> <ul style="list-style-type: none"> • Auditoría Superior de la Federación (ASF) • Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) 	<i>Instancia gubernamental evaluada</i> <ul style="list-style-type: none"> • Dependencia de gobierno
Relación gubernamental	
<i>Instancia gubernamental</i> <ul style="list-style-type: none"> • Programas de dependencias federales, estatales o municipales 	<i>Receptor/Beneficiario de proyecto</i> <ul style="list-style-type: none"> • OSC • Universidad Pública • Persona física/moral

Fuente: Elaboración propia

En un siguiente nivel, se realizan evaluaciones de impacto que basan la rendición de cuentas sobre la efectividad de los proyectos, es decir sobre la generación de resultados. En ese sentido, se realizan de forma *ex post* y requieren la medición de indicadores sensibles al cumplimiento de *outcomes*. Dado que esto requiere capacidades de evaluación específicas, normalmente son realizadas por especialistas externos a las propias organizaciones ejecutoras (Cuadro 9).

Cuadro 9. Diferencias entre los dos tipos de evaluaciones sumativas.

Objeto de evaluación	Tipo de evaluación		Grado de control del agente	Dificultad de la evaluación
Outcomes		Evaluación de impacto La rendición de cuentas es con base en los resultados obtenidos por el agente.	 Control decreciente	 Dificultad incremental
Outputs	Evaluación de desempeño La rendición de cuentas se hace con base en elementos que se asume están bajo el control del agente.			
Actividades				
Uso de Inputs				

Fuente: Modificado de Serrat, 2010

Al establecerse el nivel de desempeño esperado en planes de trabajo, contratos o convenios, las evaluaciones sumativas ejercen una influencia importante sobre la operación de los agentes. En ese contexto, el principal, es decir, aquel al que se rinde cuentas, puede usar la información proveniente de las evaluaciones para determinar si un proyecto debe continuar, terminar o ser modificado (Patton y Horton, 2009). Pero, si además del proyecto, el objeto de la evaluación es el personal o la organización con la que se colabora, entonces también podría utilizarse para tomar decisiones de contratación, promoción o continuación de una relación laboral.

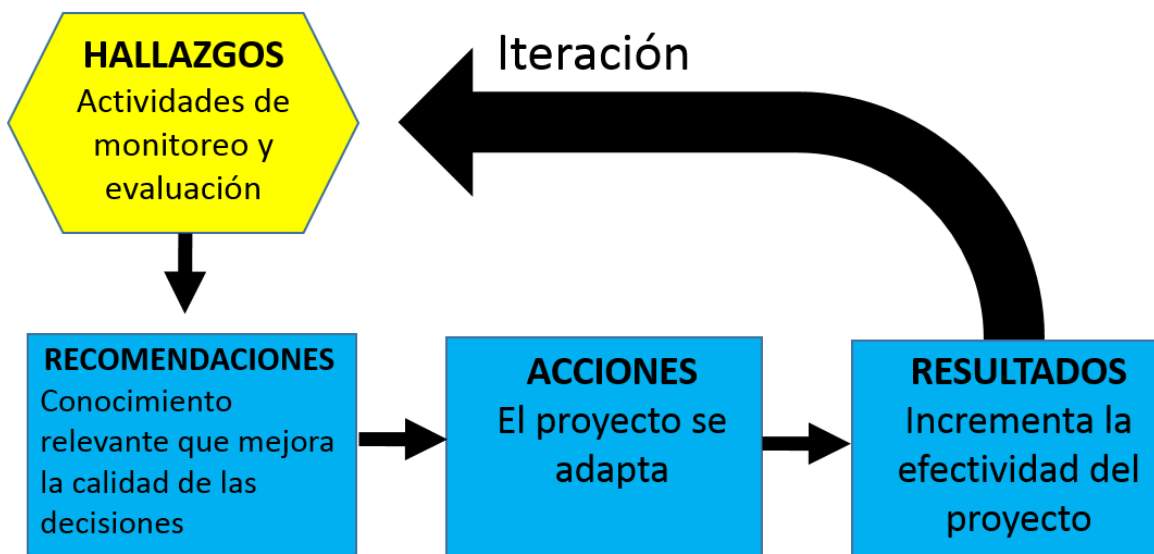
1.1.3.2 Evaluaciones formativas

En contraste, las evaluaciones formativas tienen el objetivo de alimentar ciclos de reflexión e innovación que permitan a las organizaciones mejorar la efectividad de los proyectos (Patton y Horton, 2009). Su lógica (figura 2), ligada al concepto de manejo adaptativo, es la siguiente:

Si identificas con rigor científico los resultados de lo que has hecho y los mecanismos que condujeron a esos resultados, entonces se fortalecerá tu capacidad para orientar la acción futura, y por ende, se incrementará la probabilidad de que generes efectos positivos con tu intervención (Visser *et al.*, 2013).

En otras palabras, para que los hallazgos de una evaluación incidan en los resultados de un proyecto, tienen que ser incorporados para tomar acciones concretas. Finalmente, si la lógica causal que motiva las acciones es correcta y, además, la implementación es adecuada, entonces pueden esperarse mejores resultados (Cejudo, 2011).

Figura 2. Representación de la lógica de evaluaciones formativas.



Fuente: Adaptado de Cejudo (2011) y Hansen *et al.* (2013)

El uso de evaluaciones formativas es especialmente relevante en contextos de implementación complejos, pues en estos cada situación es única y el éxito pasado provee pistas insuficientes para definir la acción futura (Snowden y Boone, 2007: 71). La incertidumbre puede reflejarse en el entorno de intervención (contextos cambiantes), en los actores involucrados (sus incentivos, redes y comportamientos) o en los mecanismos empleados por el proyecto (falta de evidencia de qué actividades funcionan mejor) (Hernández *et al.*, 2019). Por ello, dada la incapacidad para predecir lo que pasará, es crucial hacer pruebas y aprender sobre la marcha (Eyben, 2008; Guijt, 2008).

La promoción del manejo sostenible de recursos naturales se encuentra en este escenario, enfrentando un conjunto de problemas interrelacionados entre sí con altos niveles de incertidumbre y valores e intereses en competencia (Guijt, 2008). Por este motivo, no pueden “diseñarse” soluciones óptimas, sino respuestas parciales que sean suficientemente buenas (Steyaert y Jiggins, 2007; Eyben, 2014). El proceso se parece a un juego de ajedrez, en donde al mover hay que esperar reacciones sobre las que no se tiene control y en donde cada cambio de escenario se convierte en el punto de partida del siguiente movimiento¹³.

Dado el énfasis puesto en la adaptación, en este tipo de evaluación el usuario principal es el grupo responsable de la implementación. Sin embargo, para que esto pueda suceder, se requiere descentralizar la toma de decisiones al menor nivel

¹³ Scott (1998) nombró “Métis” a esta forma de razonar en escenarios complejos, la cual, confiando en la experiencia y el conocimiento práctico, privilegia tomar pasos pequeños, asumir que habrá sorpresas y permitir cambios sobre la marcha.

posible en las organizaciones, pues su autonomía es necesaria si se quiere ajustar las operaciones (Eyben, 2008; Mayne, 2010).

Además, su experiencia es una fuente de información crucial, ya que la complejidad de los problemas es más evidente para los que están en contacto directo con ellos (Chelimsky, 2015). Por este motivo, con frecuencia se privilegian evaluaciones participativas lideradas por el personal operativo, las cuales pueden tener, o no, el auxilio de consultores externos (Patton, 2010). Esto ha conducido al cuestionamiento de evaluaciones *ex post* hechas exclusivamente por evaluadores externos, pues con frecuencia sus recomendaciones llegan demasiado tarde para orientar la implementación, o solo son dirigidas a tomadores de decisión ubicados en la parte superior de la jerarquía organizacional (Mackay y Horton, 2003; Darling, 2012). En este contexto, el nuevo rol de los evaluadores profesionales es desarrollar las capacidades en evaluación del personal de las organizaciones, de forma que pongan en marcha sistemas de recolección, análisis y uso de la información proveniente de evaluaciones (King y Alkin, 2019).

Dada esta lógica, es evidente la centralidad que tiene el aprendizaje organizacional para las evaluaciones formativas. La definición clásica de Argyris y Schön (1978) indica que los integrantes de una organización aprenden para ella cuando responden a los cambios que perciben en el contexto y, como consecuencia, hacen algo mejor en el futuro. Esto implica que son capaces de detectar, compartir y corregir los supuestos erróneos de la organización. Además de tener esta función reactiva, el aprendizaje organizacional puede ser un proceso proactivo de creación de sentido (*sense-making*) que permite a la organización entender mejor el contexto y anticipar la trayectoria de situaciones con un alto grado de complejidad e incertidumbre (Guijt, 2008).

Se han distinguido tres niveles en el aprendizaje organizacional¹⁴. En la superficie se encuentra el aprendizaje de un ciclo (A1C), el cual ocurre cuando se detecta y corrige una desviación de la planeación sin cuestionar los supuestos que modelan la acción. Es decir, el aprendizaje surge al preguntarse ¿estamos haciendo las cosas bien? En caso de que no suceda lo proyectado, las modificaciones tienen la intención de poner la intervención de regreso en el camino trazado, pero sin cuestionar la validez del camino mismo.

Desde esta óptica se confía que la correcta implementación de “mejores prácticas” conduce necesariamente a los resultados esperados (Snowden y Boone 2007: 70). Por ello, lo aprendido puede conducir a procesos de capacitación cuando se detectó que la desviación fue por falta de habilidades, o bien, a una mejora operativa al corregir los errores y desatenciones identificados.

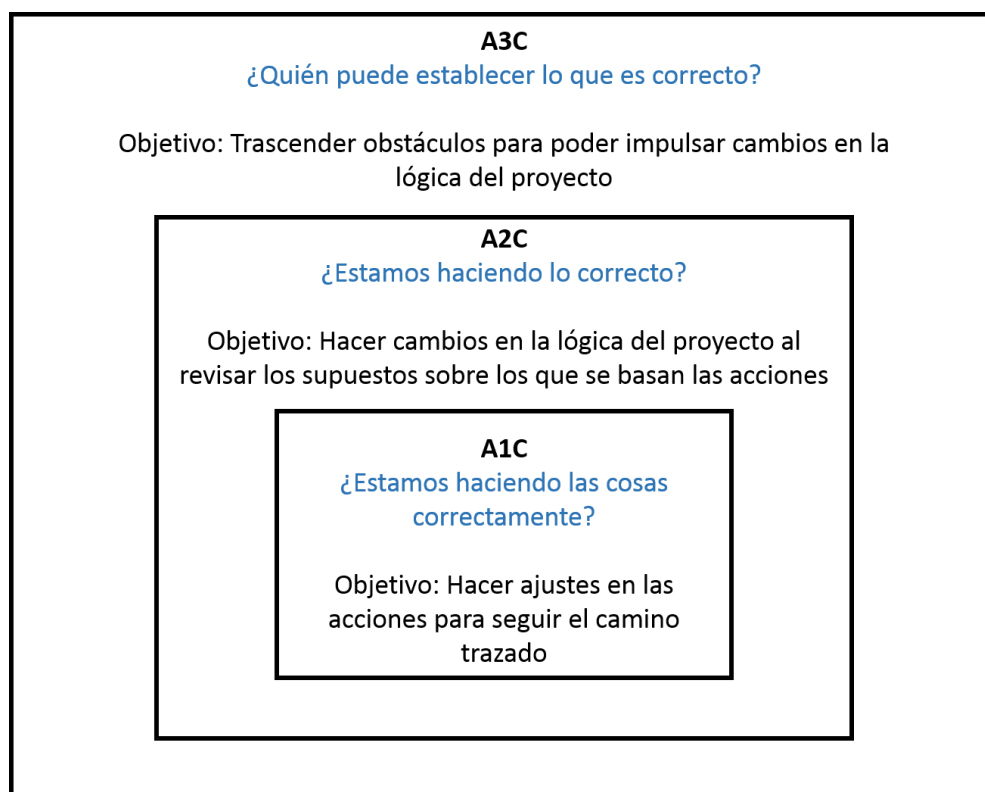
En cambio, el aprendizaje de doble ciclo (A2C) interviene tanto las acciones como los supuestos que las dirigen. A este nivel, aprender significa cuestionar los juicios implícitos en los que se basan las acciones, lo que expande el rango de opciones elegibles. El aprendizaje surge al preguntarse ¿estamos haciendo lo correcto? Hacer esta pregunta implica incorporar otros intereses y perspectivas en el

¹⁴ Los primeros dos descritos por Argyris y Schon (1996) y el tercero por Hummelbrunner (2015).

entendimiento de lo que pasa, lo que podría conducir al replanteamiento de estrategias o del propósito mismo de los proyectos.

Por último, el aprendizaje de triple ciclo (A3C) se relaciona con el análisis del contexto que determina el proceso de aprendizaje. En este sentido, involucra una reflexión crítica sobre las relaciones de poder que afectan los límites de lo que se puede cuestionar y modificar. El aprendizaje surge al preguntarse ¿Quién y cómo puede establecer lo que es correcto? Su intención es identificar los factores que constriñen la capacidad de aprendizaje para pensar cómo trascenderlos (Figura 3).

Figura 3. Niveles del aprendizaje organizacional.



Fuente: Elaboración propia a partir de Hummelbrunner (2015)

En su conjunto, los tres niveles prueban que las organizaciones solo obtienen los frutos del aprendizaje cuando combinan una actitud reflexiva y una operación flexible. Es decir, la sensibilidad para reconocer que algo no está sirviendo y el poder para modificarlo (Prieto *et al.*, 2017).

Pero, dicho lo anterior ¿cuál es la relación de las evaluaciones con esos ciclos de aprendizaje? Esa interrogante es un tópico fundamental de la teoría evaluativa. Evidentemente, la pregunta puede responderse en términos metodológicos, pero

conviene comenzar por el fundamento psicológico que vincula el acto de evaluar con virtudes epistémicas (Turri *et al.*, 2019).

Las virtudes epistémicas hacen referencia a las características propias del agente que se dispone a conocer, las cuales favorecen su adquisición de conocimientos. Algunas virtudes epistémicas son la humildad intelectual¹⁵ (Church y Samuelson, 2017), la apertura mental y la curiosidad (Turri *et al.*, 2019). En contraste, ejemplos de vicios epistémicos son el dogmatismo, el pensamiento desiderativo (*wishful thinking*) o la aversión a la crítica, rasgos que frustran las intenciones de perseguir el conocimiento (Pigliucci, 2017).

Dicho lo anterior, es innegable el vínculo entre aprendizaje y auto-evaluación. Evaluarse implica una disposición de carácter que no solo está abierta ante la posibilidad y las consecuencias del error, sino que también valora la retroalimentación que surge de ese descubrimiento. Además, pensar de forma evaluativa aporta habilidades y valores que refuerzan la capacidad para aprender, como la identificación de supuestos, la incorporación de varias perspectivas y el aprecio de la toma de decisiones informadas como preparación para la acción (Buckley *et al.*, 2015: 378).

En ese sentido, lo ideal es que las evaluaciones ayuden a trascender sesgos cognitivos que afectan la toma de decisiones (Hernández *et al.*, 2019). Algunos sesgos que pueden afectar a profesionales del desarrollo son:

- la predilección por soluciones simples y lineales que dan una sensación de control,
- la atención desproporcionada en información que confirma los supuestos que se tienen, omitiendo pruebas de su falsedad (*confirmation bias*), o
- la tendencia a continuar proyectos por el simple hecho de que ya se ha invertido en ellos, independientemente de si están funcionando o no (*sunk-cost bias*) (Banco Mundial, 2015).

Dejando de lado esta explicación psicológica, en términos prácticos se espera que las evaluaciones sean útiles para aprender al 1) formular preguntas sobre las relaciones de causalidad de un proyecto, 2) recolectar la información necesaria para darles respuesta y 3) emitir, a partir de ahí, recomendaciones de mejora. Obviamente, su efectividad depende de que la lógica causal de los cambios sugeridos sea verdadera, que los cambios se implementen de forma adecuada, y que no ocurran perturbaciones externas inesperadas (Cejudo, 2011: 15-16).

Además, se han propuesto metodologías que, más allá de proyectos, evalúan la efectividad organizacional. La intención es identificar áreas de oportunidad y proponer procedimientos que mejoren su funcionamiento. En particular, en el campo

¹⁵ La humildad intelectual permite a la persona reflexionar sobre los límites de su propio conocimiento. Gracias a ella, la persona identifica los prejuicios y supuestos de sus ideas, reconociendo los errores de su pensamiento. Esto le permite ser receptivo a diversas fuentes de evidencia y valorar la importancia de exponer sus acciones y pensamientos ante otros para recibir retroalimentación (Church y Samuelson, 2017).

de la conservación se cuenta con metodologías como el Índice de Implementación y Gestión (OLCEFS¹⁶, 2015), o la Herramienta de Seguimiento de la Efectividad del Manejo (Stolton y Dudley, 2016).

Sin embargo, un supuesto tácito de muchas metodologías de evaluación es asumir que la evidencia, una vez generada, se usará para conseguir mejores resultados en el futuro (Guijt, 2008). Se suele pensar que con la información todo caerá en su lugar, pero con frecuencia se obvia la importancia de otros procesos necesarios para transformar la información en aprendizaje, como la comunicación efectiva o el poder para tomar decisiones (Guijt, 2008).

Por ello, existe un área gris que traslapa las funciones de evaluaciones formativas con otras herramientas que sirven para tratar de entender lo que está sucediendo, como la práctica del monitoreo y múltiples técnicas para gestionar el conocimiento (Guijt, 2008; Kusters *et al.*, 2011). En ese sentido, el manejo adaptativo de un proyecto puede integrar evaluaciones con una combinación de herramientas que: 1) recolecten y organicen datos, 2) abran la oportunidad para la reflexión, 3) preparen flujos de información dirigidos a quién la necesita y 4) registren el razonamiento de las decisiones que orientaron la acción (Cuadro 10).

Cuadro 10. Algunas herramientas útiles para promover el aprendizaje organizacional en el campo del desarrollo.

Herramienta	Descripción
<i>Auditoría de conocimientos</i>	Investiga las fortalezas, debilidades y necesidades en la generación de conocimiento de una organización. Su objetivo es proponer mejoras al entender qué conocimiento se necesita, quién lo necesita, dónde se puede conseguir, y qué problemas existen.
<i>Indagación apreciativa</i>	Facilita el desarrollo de mejoras en las organizaciones. Se enfoca en lo positivo al promover la reflexión en torno de lo que está funcionando bien. La finalidad es proponer cambios que partan de los recursos y fortalezas existentes.
<i>Café Mundial</i>	Promueve la reflexión colectiva, el intercambio de conocimientos y la construcción de acuerdos a través de un proceso estructurado de conversaciones en torno de preguntas.
<i>La técnica SCAMPER</i>	Utiliza una serie de preguntas dirigidas para resolver un problema o aprovechar una oportunidad. Funciona como una técnica de lluvia de ideas que estimula la reflexividad, creatividad y flexibilidad de un equipo.
<i>Técnica de los cinco por qué</i>	Enfoque de resolución de problemas que promueve el pensamiento profundo al preguntar cinco veces cuál es la causa de una situación. El objetivo es llegar al problema de raíz.

¹⁶ Organización Latinoamericana y del Caribe de Entidades Fiscalizadoras Superiores.

Herramienta	Descripción
<i>Entrevistas de salida</i>	Mecanismo de retroalimentación para capturar el conocimiento de integrantes que están por dejar una organización.
<i>Notas de progreso mensual</i>	Mecanismo de retroalimentación que documenta los logros, problemas y cuellos de botella que se están enfrentando con la implementación de un proyecto. A su vez, solicita el registro de alternativas para lidiar con los desafíos identificados.

Fuente: Serrat, 2010

1.1.3.3 Rendición de cuentas vs aprendizaje

Algunos autores han señalado que una evaluación no puede cubrir simultáneamente propósitos de aprendizaje y de rendición de cuentas. Esto es así porque cada propósito tiene diferentes necesidades de información (Halachmi, 2002) o porque, debido a la posibilidad del sesgo de interés personal, el cual conduce a sobreestimar los beneficios generados, no es posible confiar en los hallazgos de evaluaciones formativas como un mecanismo de rendición de cuentas (Scriven, 2005). Por tanto, se asume que las evaluaciones sumativas generalmente requieren evaluadores externos, independientes y objetivos, mientras que las formativas requieren del involucramiento y apropiación de los grupos evaluados (Cuadro 11).

Cuadro 11. Comparación de evaluaciones sumativas y formativas.

	Evaluación sumativa (Rendición de cuentas)	Evaluación formativa (Aprendizaje)
Propósito central	Rendir cuentas al descubrir lo que pasó	Aprender para mejorar el desempeño futuro
Énfasis	Grado de éxito o fracaso a través de pruebas medibles	Los motivos del éxito o el fracaso, incorporando la percepción de los involucrados y partiendo de que no todo lo que cuenta puede contarse.
Usuario principal	Aquel al que se le rinden cuentas	La instancia evaluada
Estatus de la evaluación	La evaluación es un producto final	La evaluación es parte del ciclo de los proyectos
Características de los evaluadores	Deben ser independientes	Deben incluir al personal de la organización

Fuente: Elaboración propia a partir de Serrat (2010).

Lamentablemente, el financiamiento basado en proyectos y la prevalencia del paradigma de la Gestión para resultados en el Desarrollo (GpRD) han generado un desbalance a favor del uso de evaluaciones sumativas (Cousins *et al.*, 2015; Williams, 2015). En este escenario, existe el riesgo de que se simplifique el contexto de implementación de muchos proyectos, al asumir, para facilitar el control, que existe una alta predictibilidad entre las actividades y los resultados a generar. Esto ha provocado una desconexión entre la retórica de la necesidad de aprender y la presión de procedimientos que requieren el cumplimiento de objetivos preestablecidos (Guijt, 2008; Hummelbrunner, 2015).

No hay margen para el aprendizaje si se asume *ex ante* que un cambio proyectado debe ocurrir. Por ello, para evitar algunas consecuencias negativas de los “excesos” de la rendición de cuentas, como simular la obtención de resultados, se ha propuesto redirigir las evaluaciones, de forma que trasciendan la mera evaluación de proyectos y se enfoquen en la mejora del desempeño de las organizaciones (Ebrahim, 2003; Bovens *et al.*, 2008; Eyben, 2014).

Más aún, se ha dicho que no existe incompatibilidad entre aprendizaje y rendición de cuentas, pues ambos propósitos se podrían retroalimentar y cumplir simultáneamente (Alkin y Christie, 2004; Cousins *et al.*, 2015). Desde una perspectiva más amplia, la rendición de cuentas no se refiere tan sólo a la capacidad para responder a actores externos. Ésta tiene también una dimensión interna, que implica ser responsable de lo que uno ha hecho, y estar comprometido con los objetivos que se desea conseguir (Ebrahim 2003; Guijt 2010; Roetman, 2011). En el fondo, este deseo por mejorar algo en la sociedad, más que la solicitud externa de resultados, es la fuerza que debería motivar la búsqueda de logros.

En consecuencia, una organización comprometida será un aprendiz tenaz, pues buscará respuestas para resolver los problemas que le preocupan, y al mismo tiempo, generará información para responder a otros y a sí misma si los recursos se están utilizando de la manera más eficaz. Inclusive, este enfoque amplía los alcances de la rendición de cuentas, pues involucra la opinión de los grupos involucrados y solicita evidencias de las adaptaciones que se han hecho (Cuadro 12).

Desde esta perspectiva, un número creciente de donantes se están convirtiendo en aliadas de sus donatarios, pues están transitando de un modelo de rendición de cuentas basado en la exigencia de resultados de corto plazo, a uno en el que se apoyan los esfuerzos de las organizaciones por adaptarse y aprender dentro de un contexto que cambia rápidamente (Eyben, 2008; Darling, 2012).

Ese tipo de relación implica que los donantes descendan del terreno elevado que establece lo que “debería ser”, para acompañar a las organizaciones en el terreno pantanoso donde es evidente que los problemas son confusos, las elecciones implican dilemas y existen barreras institucionales que dificultan la adopción lineal de buenas ideas (Schön, 1995). En este contexto, en lugar de soluciones perfectas,

se buscan soluciones satisfactorias que permitan ir saliendo adelante¹⁷ (Mackay y Horton, 2003).

Cuadro 12. Transición hacia una rendición de cuentas transformacional.

Rendición de cuentas	Proyectos
No transformacional	Usar la evaluación para revisar que los recursos se utilizaron como se había comprometido y para saber si las actividades produjeron o no los efectos esperados.
Transformacional	Usar la evaluación para mejorar la efectividad de un proyecto y, desde ahí, rendir cuentas a los grupos involucrados. Es decir, demostrar el grado en el que se tomaron decisiones basadas en lo aprendido, y cómo esto permitió responder mejor al contexto de la intervención.

Fuente: Elaboración propia a partir de Cousins *et al.* (2015).

Esto no quiere decir que desaparezcan las evaluaciones de desempeño, pues no todo es impredecible en la operación de proyectos de desarrollo. Algunos procesos pueden evaluarse en función del cumplimiento de metas preestablecidas y otros con base en el aprendizaje de los errores (Guijt, 2010; Eyben, 2014).

Sin embargo, actualmente siguen vigentes normas e incentivos que dificultan transitar hacia este modelo de gestión, y por ello, es muy difícil encontrar un balance adecuado para combinar lo mejor de ambos propósitos de evaluación. Un ejemplo se encuentra en la falacia de la falla, la cual externa que, más allá de visiones románticas, aprender a partir de los errores propios sólo se celebrará en la medida en la que posteriormente se tenga éxito (Nussbaum, 2013).

Un punto clave está en el tiempo que se otorga para lograr cambios transformacionales. Así como en el mundo de las finanzas se acepta que la obtención de retornos ocurra en plazos superiores a cinco años (Graham, 2006), en el campo del desarrollo es importante redefinir la percepción del éxito e invertir en procesos de largo plazo que valoren la obtención de pequeños cambios incrementales (Kessler, 2006; Guijt, 2008).

¹⁷ Traducción libre de la expresión *muddle through*.

1.2. LAS EVALUACIONES COMO PROCESO

Decision making is a prime arena for developing and enjoying an interpretation of life and one's position in it. As a result, the link between the collection and display of information and its use in affecting decisions is weak.

–James March–

A pesar de la popularidad de las evaluaciones, es erróneo asumir que, una vez realizadas, éstas serán utilizadas para mejorar el desempeño de una organización. El malestar que provocó el descubrimiento de esta disonancia llevó a que se tematizara el uso de evaluaciones como objeto de estudio (Leviton y Hughes, 1981; Cousins y Leithwood 1986). La utilidad es esencial, pues las evaluaciones sólo contribuyen al aprendizaje de una organización cuando conducen a cambios en su comportamiento (Ebrahim, 2005). Es decir, el valor de una evaluación no está en su producción *per se*, sino en el valor ocasionado por su uso, al informar decisiones, mejorar la comprensión o contribuir al aprendizaje (Feinstein, 2001).

Los primeros llamados a explorar el tema se hicieron a inicios de la década de los setenta, cuando se señaló que existían muy pocas pruebas de que las investigaciones sociales y las evaluaciones impactaban en las decisiones de los gobernantes (Alkin y King, 2016). Por ello, la agenda inicial de investigación tuvo el propósito de entender mejor cómo, cuándo y por qué se usaban las evaluaciones (Weiss, 1972 en Vo, 2015). Actualmente es el área más investigada en el campo de las evaluaciones (Christie, 2007 en King y Stevahn 2013), incluyendo los llamados a abrir la caja negra y observar cómo se utiliza evidencia para tomar decisiones adaptativas (Hernández *et al.*, 2019)

Es tal la relevancia puesta en el tema de la utilidad, que se le considera uno de los cinco estándares internacionales de la evaluación de programas (Yarbrough *et al.*, 2011). Estos estándares, que abordan aspectos vinculados con la calidad de las evaluaciones, buscan guiar la forma en que éstas se llevan a cabo (Cuadro 13).

En el estudio del uso, una conclusión fructífera ha sido el reconocimiento de que las evaluaciones, lejos de ser la simple aplicación de una metodología para establecer causalidad, son un proceso social que está inmerso en el diverso y conflictivo contexto de la toma de decisiones (Weiss, 1999; Agudo, 2015). Gracias a esta comprensión, se comenzó a valorar las evaluaciones por sus resultados, es decir, por su capacidad para habilitar canales de conocimiento que mejoren la práctica de las organizaciones (Sanderson, 2000).

Cuadro 13. Estándares internacionales de la evaluación de programas.

Estándares	Descripción
1. Utilización	Busca asegurar que la evaluación servirá a las necesidades de los que participan en torno del programa evaluado. Para ello, es fundamental que estos consideren que los procesos y productos de la evaluación son valiosos.
2. Factibilidad	Busca asegurar que la evaluación será realista, siguiendo criterios de eficiencia y eficacia.
3. Propiedad	Busca establecer que la evaluación se conducirá de forma legal y ética, de forma que se respete a los involucrados en la evaluación
4. Exactitud	Tiene el propósito de asegurar que una evaluación aportará información veraz de los aspectos que determinan valor o mérito en el programa que está siendo evaluado.
5. Rendición de cuentas	Promueve que exista una documentación sobre la forma en que se hizo la evaluación, con la finalidad de mejorar procesos y productos de la evaluación.

Fuente: Yarbrough *et al.* (2011).

Es importante hacer una distinción clave en el estudio del fenómeno. Por un lado, se ha investigado el uso de evaluaciones externas, es decir, cuando la carga del desarrollo de la evaluación está fuera de la organización evaluada, pero el usuario de las recomendaciones es la organización evaluada o un tercero (Cejudo y Maldonado, 2011; Cardozo, 2015). Por el otro, se ha examinado el uso de evaluaciones internas, contexto en el que las organizaciones hacen por sí mismas las evaluaciones (Taylor *et al.*, 2013; Lennie *et al.*, 2015).

En cualquiera de los dos casos, el uso ocurre cuando los hallazgos y/o recomendaciones son tomados en cuenta por personas que tienen la capacidad para actuar con base en estos (Stake, 1986 en Alkin y King, 2017). Con el paso de los años, se ha construido una taxonomía para clasificar las distintas formas en las que pueden usarse las evaluaciones. Este aparato conceptual para estudiar el fenómeno, a su vez, ha permitido el desarrollo de investigaciones para identificar los factores que contribuyen o inhiben esos tipos de uso (Vo, 2015). En los siguientes apartados se describirá con mayor detalle cada una de estas áreas.

1.2.1 Taxonomía del uso de evaluaciones

En este apartado se presentan algunas categorías que se han propuesto para caracterizar los distintos tipos de uso. Los primeros estudios comenzaron con una comprensión limitada, al concentrarse en el uso instrumental de los resultados para informar la toma de decisiones sobre si un programa debía terminar, continuar o modificarse (Nunneley *et al.*, 2015). Desde esta óptica, basada en teorías orientadas hacia la decisión, era crítico generar información para asistir a administradores clave en la toma de decisiones, normalmente aquellos que encomendaban la evaluación de políticas públicas (Alkin y Christie, 2004).

Después, se describieron otros dos tipos de uso que capturan situaciones que no podía discernir el uso instrumental: el uso simbólico y el uso conceptual (Leviton y Hughes, 1981). Juntos, han sido la clasificación clásica de las formas en que se pueden usar las evaluaciones. Sin embargo, dada la gran diversidad de formas en las que puede ser usada una evaluación, no existe en la literatura una taxonomía estandarizada (Cardozo, 2015). A continuación, se describen los tres tipos de uso mencionados, sumando a ellos el uso externo y del proceso de evaluaciones (Cuadro 14).

Cuadro 14. Descripción de cinco tipos de uso de evaluaciones.

Propósito	Descripción
Instrumental	La evaluación conduce a acciones inmediatas y específicas, como la continuación de un programa, su revisión o su cancelación.
Conceptual	Se relaciona con una mejora en el entendimiento del problema y sus posibles soluciones, sin implicar, por ello, la toma de una acción.
Simbólico	Incluye acciones legítimas, como el uso de evaluaciones para posicionarse como una organización exitosa, e ilegítimas al usar evaluaciones para justificar posiciones pre-existentes.
Externo	Cuando la evaluación conduce a cambios más allá de lo evaluado directamente.
Del proceso	Cuando, más allá de sus recomendaciones, participar en el proceso de una evaluación promueve cambios a nivel individual y organizacional y facilita que se retome el conocimiento para tomar decisiones.

Fuente: Kusters *et al.* (2011); Alkin y King (2016).

Uso instrumental. Se presenta cuando la información de la evaluación es utilizada por tomadores de decisión para juzgar el mérito de un proyecto o programa, y por ende, para determinar si éste debe continuar, modificarse o suspenderse (Leviton y Hughes, 1981). De forma más amplia, se refiere a las circunstancias donde los

hallazgos de la evaluación se traducen directamente en el mejoramiento de un proyecto o programa, normalmente a través de la modificación de acciones y procesos para incrementar su efectividad y eficiencia (Leviton y Hughes, 1981). En este sentido, es el tipo de uso más fácil de identificar, pues implica aportar evidencia de cómo el conocimiento condujo a cambios específicos.

Uso conceptual. Esta categoría cubre aquellas situaciones donde los hallazgos de una evaluación facilitan que los involucrados en proyectos mejoren su comprensión del problema que buscan atender, sin que, por ende, se empleen de una manera específica y documentada (Weiss, 1982). La información, en cambio, sirve al decisor para ver las cosas de forma diferente, justificar ideas para hacer reformas, o sugerir la necesidad de un cambio (Weiss, 1977).

Su medición es, por tanto, compleja pues se relaciona con cambios en la forma de pensar relaciones y supuestos de los proyectos. Por ello, se ha aseverado que el uso conceptual de las evaluaciones es más frecuente de lo que puede percibirse y que tiene un impacto significativo sobre el diseño y ajuste de políticas públicas (Weiss *et al.*, 2005). Dada su vocación hacia el cuestionamiento de los supuestos que soportan las acciones, es considerada la categoría clásica más cercana al concepto de aprendizaje (Cousins *et al.*, 2015).

Uso simbólico. En el campo del desarrollo, los actores basan su credibilidad a través del uso de procedimientos legitimados socialmente, como las evaluaciones (Ebrahim, 2005; Okail, 2009). Por ello, las organizaciones pueden apoyarse en éstas, no tanto por la información que generan, sino por su función simbólica, la cual les permite manejar la forma en que son percibidas. De una forma legítima, las organizaciones pueden emplear las evaluaciones como una ventaja competitiva en la gestión de recursos, pues son un medio para comunicar que son exitosas (Preskill y Boyle, 2008; Lennie *et al.*, 2015).

Sin embargo, en un sentido ilegítimo el uso simbólico ocurre cuando se utiliza la evaluación para justificar posiciones preexistentes (Leviton y Hughes, 1981). Es decir, los hallazgos pueden usarse de forma selectiva o distorsionarse para soportar una decisión o demostrar un resultado (Agudo, 2012; Mosse, 2004). Este tipo de uso está asociado, por tanto, con las relaciones de rendición de cuentas, principalmente cuando se percibe que podría haber sanciones en caso de no cumplir con expectativas de desempeño (Ebrahim, 2005).

Como se ha dicho, en proyectos complejos los actores tienen poco control sobre los eventos. Por ello, si las desviaciones de objetivos preestablecidos son sancionadas, es más práctico controlar la interpretación de los eventos. En este contexto, las evaluaciones son un artefacto idóneo para ello (Mosse, 2004). Agudo (2015) pone un buen ejemplo de esto, al criticar el uso autorreferencial de evaluaciones externas como un medio para probar la validez de la teoría del cambio de un programa de gobierno.

...los programas de política social necesitan, para su propia puesta en funcionamiento, de legitimidad y apoyos conseguidos mediante modelos convincentes, para lo cual los diseñadores, expertos y evaluadores se afanan en la organización de interpretaciones sobre impactos planificados (p. 27)... la

utilidad de las evaluaciones no consiste en afirmar lo que ya todos sabemos, sino en apuntalar lo que queremos saber todos los que tenemos diversos intereses y participaciones en aquéllas (p. 203).

Por otra parte, se ha documentado el riesgo del uso ritual de evaluaciones que valoran *ex ante* el impacto social de proyectos de desarrollo. Este tipo de evaluación requiere la participación de los actores locales en el proceso de identificación de los impactos positivos y negativos que podría generar un proyecto de desarrollo. Adicionalmente, solicita el diseño participativo de medidas para evitar, reducir, mitigar o compensar posibles efectos negativos. Sin embargo, se ha señalado que algunas evaluaciones se llevan a cabo como un ejercicio performativo que permite aparentar la inclusión de salvaguardas sociales, sin que en realidad se les transfiera poder de decisión o autoridad a las comunidades afectadas (Mackay, 2005).

Es importante distinguir que el uso simbólico no implica necesariamente que la organización evaluada no esté interesada en los objetivos de sus proyectos. Por ejemplo, una organización comprometida con el cambio y no con la simulación, pero presionada para comportarse contra su buen juicio por requerimientos preestablecidos, puede cumplir cínicamente con una evaluación para asegurar el financiamiento mientras hace en la medida de lo posible lo que le parece más apropiado (Eyben, 2014).

Lo que es un hecho es que desde esta óptica, las evaluaciones se basan en la necesidad de mantener relaciones, y no en mejorar los proyectos ni incidir en las políticas públicas. En este sentido, las evaluaciones de proyectos se convierten en sellos de aprobación con la finalidad de garantizar el mantenimiento del flujo de recursos o la credibilidad de las organizaciones (Mosse, 2004; Agudo, 2015). Lamentablemente, con frecuencia este tipo de uso permite obtener beneficios privados a costa de los que obtendría la sociedad si las evaluaciones condujeran hacia mejoras en el manejo de los proyectos (Shulha y Cousins, 1997).

Uso externo. Esta categoría, que incorpora ideas vinculadas al concepto de influencia propuesto por Kirkhart (2000), explica situaciones en las que una evaluación conduce a cambios en ámbitos más allá de lo evaluado directamente (Kusters *et al.*, 2011). En ese sentido, implica retomar los hallazgos de evaluaciones realizadas a proyectos semejantes para persuadir sobre la importancia de realizar una determinada acción, como modificar el diseño de una política pública o adecuar las actividades de un proyecto. Por ejemplo, podría ser que una organización que opera proyectos de manejo integrado de cuencas en México utilice los hallazgos de evaluaciones hechas a proyectos similares en Perú para orientar su propia práctica, o bien, que un gobierno estatal retome los resultados de evaluaciones hechas a programas de atención a la pobreza en otro estado para rediseñar sus propios programas.

Uso del proceso de una evaluación. Desde las categorías clásicas sobre el uso de la evaluación (instrumental, conceptual, y simbólica) es común que se considere que los usuarios de las evaluaciones son funcionarios de gobierno de primer nivel, los cuales tomarán decisiones en función de los hallazgos. Sin embargo, Michael Patton revolucionó la teoría sobre el uso de las evaluaciones al dejar de enfocarse

en las necesidades de tomadores de decisión, sino en procedimientos que logran incrementar el uso de las evaluaciones entre un número mayor de usuarios potenciales (Alkin y Christie, 2004).

El enfoque de la evaluación orientada hacia el uso afirma que es más probable que las personas usen los resultados de una evaluación si participan y se apropian del proceso de la evaluación. De esta manera, el uso del proceso de evaluación

“es indicado por cambios individuales en la forma de pensar y comportarse, y cambios organizacionales en procedimientos y cultura, que ocurre entre los que están involucrados en evaluaciones como resultado del aprendizaje que ocurre durante el proceso de la evaluación (Patton, 1997: 90).

El supuesto subyacente es simple: la probabilidad de que una evaluación se use, se potencia con la identificación de personas que tienen un interés personal en los hallazgos que ésta genera. A esto Patton le llama el factor personal, y es vital para los evaluadores obtenerlo al involucrar a los usuarios en el proceso de evaluación, no como proveedores de información sino en la toma de decisiones sobre su diseño, implementación y uso (Alkin y Christie, 2004).

El uso del proceso tiene varias consecuencias en una organización. Al desarrollar su capacidad para hacer evaluaciones, las organizaciones se sensibilizan sobre la importancia de generar conocimiento, recolectar información, y cuestionar los supuestos básicos de la organización, desde las actividades, hasta las estrategias y objetivos (Cousins *et al.*, 2014).

La intención explícita de incidir en las organizaciones, a través del proceso de una evaluación, influyó en el surgimiento de enfoques de evaluación participativos orientados hacia el desarrollo de las capacidades de evaluación en las organizaciones (Fetterman, 1996; Preskill y Torres, 1999; King y Stevahn, 2013; Cousins *et al.*, 2014). Estos autores sugieren que el aprendizaje organizacional es resultado de la participación de individuos en el proceso de una evaluación colaborativa (Preskill y Torres, 2000).

Las evaluaciones que se conducen bajo este enfoque, de corte formativo, buscan que “el evaluador y el personal de las organizaciones participen y colaboren de manera conjunta en el diseño particular de la evaluación, en la recopilación de la información y en la discusión de los cambios necesarios (Pérez, 2015: 5).

1.2.2 Factores que influyen en la producción y uso de evaluaciones

Diversas investigaciones han identificado una serie de factores que inciden en la producción y uso de evaluaciones. Destaca la relevancia de variables personales, organizativas e institucionales que se entrelazan entre sí (Cousins y Leithwood, 1986; Shulha y Cousins, 1997; Cardozo, 2015; González, 2016). Sin embargo, dado que su estudio es complejo y contextual, no existe un consenso sobre el grado de influencia que ejercen (Cardozo, 2015). Además, es importante entender que las variables pueden ser positivas, al facilitar el uso de evaluaciones para tomar

decisiones, o negativas, al obstaculizarlo. Por ello, un análisis integral debería estudiar la presencia e intersección de ambas.

Se elaboró un marco conceptual escalonado para sistematizar los factores señalados en la literatura y jerarquizar su influencia (Figura 4). En primera instancia, se encuentran variables relacionadas con la **calidad de las evaluaciones**. En esta escala, centrada en aspectos metodológicos, el uso se explica por la disponibilidad de información confiable para tomar decisiones cuando se requiere (Alkin, 1985). Esto implica disponer de conocimientos técnicos que garanticen la calidad de la información¹⁸ (Cousins *et al.*, 2014), el diseño de proyectos evaluables y el fomento de procesos colectivos de aprendizaje. Para abordar estas variables, el objeto de estudio más visible es el artefacto mismo, es decir, los documentos que a) plasman la lógica interna de los proyectos, b) describen metodologías de evaluación o c) reportan los resultados obtenidos. A través de ellos es posible reconocer el propósito y características técnicas de las evaluaciones.

Figura 4. Factores que afectan la producción y uso de las evaluaciones.



Fuente: Elaboración propia.

En el siguiente nivel, influyendo sobre las variables metodológicas, se encuentran **factores organizacionales**. Después de todo, aunque la información es importante, las decisiones no se toman en el vacío. Las organizaciones son el usuario principal de las evaluaciones y por ello es importante entender si cuentan con procedimientos e incentivos para hacerlas y retomar sus resultados. La práctica evaluativa requiere el desarrollo de puestos de trabajo, sistemas y planes que integren las evaluaciones dentro de la forma en que la organización cumple con sus objetivos (Preskill y Boyle, 2008, p. 444). Además, son fundamentales la disponibilidad de recursos (tiempo y dinero) y un ambiente organizacional que favorezca el reporte de fallas y la transferencia de lo aprendido al trabajo cotidiano (Cousins *et al.*, 2014).

¹⁸ Por ejemplo, elegir la metodología en función del propósito, seleccionar indicadores pertinentes, o saber cómo recolectar datos.

En última instancia, las organizaciones operan guiadas por **arreglos institucionales** que influyen en su comportamiento. De esta manera, por mandato o por su dependencia hacia recursos externos, sus evaluaciones pueden seguir parámetros y propósitos no definidos por ellas mismas (Guijt, 2008). Aun así, es importante identificar el margen de maniobra de las organizaciones para adoptar o no los procedimientos, así como para desarrollar esquemas de evaluación siguiendo sus propios propósitos.

Debido a que algunas personas utilizan las palabras organización e institución como sinónimos, es necesario hacer una precisión conceptual. Por instituciones, se hace referencia a patrones de comportamiento establecidos, los cuales estructuran las interacciones sociales de los actores (Bourdieu, 1990). En un sentido general, esto incluye creencias, valores y teorías paradigmáticos que establecen las “reglas del juego” desde las que operan los actores. Algunos arreglos institucionales son formales, como normas y leyes, pero otras son informales como costumbres sociales. Mientras tanto, las organizaciones son sistemas de acción colectiva en donde grupos organizados de personas trabajan juntas por un propósito común (Czarniawska, 2012). Para conseguirlo, estas entidades cuentan con reglas y procedimientos que dictan el comportamiento de los subgrupos de trabajo que desempeñan funciones especializadas. En resumen, las organizaciones operan en el marco de determinados arreglos institucionales, los cuales tienen influencia sobre lo que puede ser hecho en ellas.

A continuación, se hace una descripción detallada de los tres niveles.

1.2.2.1 Calidad de las evaluaciones

En última instancia, la calidad de una evaluación se expresa cuando se usa para lo que fue hecha. Cuando se trata de evaluaciones externas, si el usuario percibe que la evaluación aporta información confiable, y además está disponible justo en el momento en el que la necesita, entonces por sus características mismas la evaluación facilitará la toma de decisiones (Cousins y Leithwood, 1986; Kusters *et al.*, 2011). Por el contrario, si los usuarios no perciben que la información es creíble, si la información no se encuentra a la mano cuando las decisiones son tomadas, o si los hallazgos son irrelevantes, pues no ayudan a resolver los problemas que le interesaban a los usuarios de la evaluación; entonces, con certeza, la evaluación no será retomada para propósitos legítimos¹⁹ (Feinstein, 2001; González, 2016; Cejudo y Abarca, 2016).

Por su parte, cuando la organización es a un tiempo productora y usuaria de sus evaluaciones, además, es crucial que personal de la organización tenga la capacidad técnica para a) diseñar proyectos evaluables (Peersman *et al.*, 2015), b) desarrollar evaluaciones (Cousins *et al.*, 2014) y c) fomentar procesos colectivos de aprendizaje a partir de ellas (Roetman, 2011).

¹⁹ Una evaluación de mala calidad podría utilizarse, de cualquier manera, de forma simbólica.

a) Diseñar proyectos evaluables

La evaluabilidad hace referencia al grado en el que un proyecto puede ser evaluado de manera confiable. Para hacer una evaluación que establezca la causalidad entre actividades y efectos, es necesario, en primer lugar, que sea plausible que las actividades conduzcan a los efectos deseados, que los resultados a alcanzar se enuncien de manera clara y que sea factible recopilar información para demostrar la causalidad. Muchas veces no se toman en cuenta estos aspectos, por lo que una buena parte de los problemas de evaluación se originan desde el diseño de las intervenciones (Peersman *et al.*, 2015).

En ese sentido, un primer paso esencial para hacer evaluaciones es desarrollar la teoría sobre cómo se supone que el proyecto funcionará (Bickman, 1987; Weiss, 1997). Al hacer una teoría del cambio se vinculan los recursos, actividades y resultados de un proyecto en una cadena de supuestos causales (Funnel y Rogers, 2011), y, gracias a esa representación, es posible cumplir con múltiples objetivos de la evaluación de proyectos (Cuadro 15). En el campo de la conservación se ha señalado su importancia para clarificar objetivos, definir indicadores de desempeño, y desarrollar las hipótesis que se pondrán a prueba (Roe *et al.*, 2013; Margoluis *et al.*, 2013; Baylis *et al.*, 2016; Game *et al.*, 2018; Salafsky *et al.*, 2019).

Cuadro 15. Contribuciones de la teoría del cambio (TdC) a la evaluación de proyectos.

Rubro	Descripción
<i>Planeación estratégica</i>	La TdC promueve la clarificación tanto de objetivos como de estrategias para conseguirlos (Pasanen <i>et al.</i> , 2019).
<i>Desarrollo de indicadores</i>	La TdC es un marco útil para elegir los indicadores con los que se medirá el progreso en la ruta de cambio esperada (Richards y Panfil, 2011).
<i>Prueba de hipótesis</i>	La TdC fuerza a que se explicita el conjunto de supuestos en los que se basa la lógica de un proyecto. Gracias a ello, es mucho más fácil identificar las hipótesis que serán puestas a prueba durante la fase de implementación (Pasanen <i>et al.</i> , 2019).
<i>Manejo adaptativo</i>	Lo más probable es que varios supuestos sobre eventos críticos no se sostendrán, por lo que visitar la TdC sirve para promover la reflexión en torno de las adaptaciones requeridas, tanto en la implementación como en las preguntas para medir correctamente lo que ocurre con un proyecto (Hernández <i>et al.</i> 2019; Pasanen <i>et al.</i> , 2019).

Fuente: Elaboración propia

b) Desarrollar evaluaciones

Por su parte, desarrollar evaluaciones implica contar con un conjunto de conocimientos y habilidades, las cuales son requeridas en distintas fases del proceso (Cousins *et al.*, 2014). En primera instancia, para planificar una evaluación se tiene que saber cuáles son los requerimientos de aprendizaje y rendición de cuentas, pues a partir de ellos se esclarecerán las necesidades de información. Con base en lo anterior, así como en las características del proyecto a evaluar y los recursos disponibles, se elegirán los métodos más apropiados y se diseñarán los instrumentos de recolección de información (Chaplowe y Cousins, 2016; Pasanen *et al.*, 2019).

Después, es evidente que se requiere rigurosidad en la recolección y análisis de información. Esto implica saber triangular con otras fuentes de información y perspectivas cuando la evidencia es incompleta (Van Hemelrijck y Guijt, 2016), pero también reconocer que no siempre se necesitan los mismos estándares en la generación de evidencia para tomar una decisión (Salafsky *et al.*, 2019).

c) Uso del proceso

El involucramiento en las evaluaciones de los que están interesados en el funcionamiento de un proyecto es uno de los factores que mayor influencia tiene en el uso de evaluaciones (Cousins y Leithwood, 1986; Shulha y Cousins, 1997; Weiss, 1998; Preskill y Torres, 2000; Cousins *et al.*, 2004; Johnson *et al.*, 2009; Kusters *et al.*, 2011; Cardozo, 2015; Pérez, 2015; Cousins *et al.*, 2015).

Por ello, se han señalado las limitaciones de evaluaciones externas con poca participación de los usuarios, pues con frecuencia las recomendaciones no sólo llegan demasiado tarde para guiar la implementación, sino que sitúan el aprendizaje en los hombros de evaluadores externos, lo que hace que el ciclo de aprendizaje sea menos eficiente (Darling, 2012).

Si es verdad que la apropiación de la evaluación favorece el uso de los hallazgos, como establece el uso del proceso (Patton, 1997), entonces la participación de los grupos involucrados debería garantizarse en el diseño metodológico de la evaluación, y por tanto debería ser un aspecto vinculado con su calidad.

Cousins y otros autores (2015) argumentan que el proceso de aprendizaje es un aspecto central para el uso, pues crea el espacio para él. Si se experimenta una evaluación exitosa porque fue útil, entonces esto conducirá a que se valore la importancia de las evaluaciones. En ese sentido, concluyen, experimentar el éxito en el uso de evaluaciones es el pivote de su sustentabilidad. Por otro lado, entender la evaluación como un proceso de aprendizaje facilita la construcción de confianza entre los participantes de una evaluación, pues ésta deja de ser asociada con una tarea estresante cuya única función es juzgar el desempeño (Cousins *et al.*, 2015).

Esto implica, en buena medida, dejar de juzgar la calidad de las evaluaciones exclusivamente en función de la sofisticación de las metodologías para establecer la causalidad, e incrementar el valor de una evaluación en función de su capacidad

para habilitar canales de conocimiento que mejoren la práctica de las organizaciones (Guijt, 2008). En este sentido, el personal de la organización que sea responsable de evaluaciones debe convertirse en un facilitador que sepa guiar el proceso de aprendizaje (Alkin, 1985; Preskill y Torres, 2000), y por ello, requiere dominar un conjunto de prácticas que le permitan promover la reflexión y la adaptación (Guijt, 2008).

1.2.2.2 Variables organizacionales

Disponer de evidencia es una condición necesaria, más no suficiente para usar evaluaciones. Dado que el uso ocurre en el contexto de una organización, se requiere poder tomar decisiones, estar interesado en promover mejoras y disponer de recursos para implementar cambios. Parece obvio, pero estos elementos con frecuencia no se encuentran presentes de forma simultánea en las organizaciones (Cejudo y Abarca, s.f.: 2). Por ello, la literatura sobre el uso resalta la importancia de entender qué tan preparadas están las organizaciones para hacer evaluaciones y para aprender a partir de ellas (Vo, 2015).

Las evaluaciones son una parte del sistema organizacional de aprendizaje, por tanto, en el fondo su uso depende de la capacidad y compromiso que tengan las organizaciones para aprender. Si bien es necesario que sus integrantes sepan diseñar proyectos evaluables y preparar evaluaciones de calidad, los procedimientos organizacionales dan el motivo, los medios y la oportunidad para que las evaluaciones ocurran (Serrat, 2010). Es decir, incluso cuando los integrantes de una organización tienen el conocimiento y la motivación por adaptar su práctica, si no existe un ambiente organizacional que favorezca que esto ocurra, entonces las evaluaciones generarán, en el mejor de los casos, pequeños espacios de aprendizaje (Taylor *et al.*, 2013; Cousins *et al.*, 2014; Taut, 2007).

Por tanto, para que las prácticas de evaluación se puedan sostener, los integrantes precisan contar con el soporte de los líderes, incentivos y recursos para transferir lo aprendido sobre evaluaciones a su trabajo cotidiano. La práctica evaluativa sostenida también requiere el desarrollo de sistemas, procesos, y planes que ayudan a integrar las evaluaciones en la forma en que la organización cumple con su misión y con sus objetivos estratégicos (Preskill y Boyle 2008: 444).

A continuación, se desarrollan una serie de variables organizacionales ligadas, positiva o negativamente, con el uso de evaluaciones.

a) Cultura organizacional

A diferencia de las personas, las organizaciones no piensan o actúan (Buckley *et al.*, 2015). Por lo tanto, una organización que aprende puede ser descrita mejor como una organización en la que las personas, a todos sus niveles, valoran y se involucran en actividades de aprendizaje.

Promover un comportamiento colectivo requiere modelos mentales que favorezcan, a través de un lenguaje común, la integración de las personas que conforman una

organización (Ekboir *et al.*, 2009). En ese sentido, la cultura de una organización reúne los valores y supuestos subyacentes que rigen su accionar (Grantmakers for Effective Organizations, 2016).

Los **valores** de la organización expresan lo que la organización espera conseguir y cómo considera que es posible hacerlo (GEO, 2016). A través de distintos artefactos, las organizaciones comunican de forma explícita su creencia en ciertos conceptos esenciales, por ejemplo, manejar a través de adaptaciones. En teoría, esos valores estructuran la forma en la que trabajan, pero podrían no verse manifestados de forma tangible en la estructura organizacional y los procedimientos de la organización. Esto es así, pues los **supuestos subyacentes** son los principios que dirigen realmente el trabajo en una organización. Estos supuestos pueden seguir los valores de la organización, o ir en contra de ellos (GEO, 2016).

El fenómeno ha sido descrito por Argyris y Schön a través de la distinción entre dos teorías de la acción (1978): la teoría adoptada y la teoría en uso. La primera hace referencia a las palabras que usa una organización para expresar lo que hace y por qué, o también, lo que quisiera que otros pensaran que hace. La teoría en uso, en cambio, son las ideas que están detrás de las acciones que realiza la organización. Mientras más congruencia haya entre la teoría adoptada y la teoría en uso “habrá más efectividad, en el sentido de que exista una divergencia menor entre expectativas y consecuencias” (Guijt, 2008: 109).

Por ejemplo, una organización podría manifestar públicamente que las evaluaciones sirven para aprender (teoría adoptada), pero aun así no llevar a cabo actividades para que esa creencia se convierta en una realidad porque en realidad se concibe a las evaluaciones como un mecanismo para juzgar y sancionar (teoría en uso).

Argyris (1994) explica la disociación por rutinas defensivas tan internalizadas que es difícil para los integrantes de la organización identificar y cuestionar las incongruencias. Más aún, incluso si el personal reconoce la brecha entre las acciones y los valores de la organización, su toma de decisiones puede estar influenciada por su percepción de las relaciones de poder en su entorno organizacional. Su comprensión de las “reglas del juego”, adquirida a través de la experiencia, le permite discernir entre lo que debe acatar y lo que puede cuestionar (Okail, 2009).

Dicho lo anterior, una organización con una cultura evaluativa en uso daría muestras de ello al²⁰:

- Reconocer los límites de su propio conocimiento y valorar la importancia de exponer sus acciones y pensamientos ante otros para recibir retroalimentación.
- Buscar evidencia de lo que está sucediendo para cuestionar y soportar sus decisiones
- Valorar el diálogo genuino y abierto.

²⁰ Las ideas parten de Mayne (2010).

- Hacer tiempo para aprender.
- Aprender de los errores y del mal desempeño.
- Promover la transferencia de conocimiento.
- Apoyar la toma de riesgos.
- Buscar nuevas maneras de resolver los problemas.

Evidentemente, es sumamente compleja la medición del uso de estos valores organizacionales. Por ello, se puede rastrear su expresión en manifestaciones visibles del comportamiento organizacional como su **estructura operativa** y sus **sistemas de soporte** (Serrat, 2010).

b) Estructura operacional

La estructura operacional de una organización permite comprender el tamaño y nivel de especialización de una organización, pues en ella se expresa la distribución de funciones, la coordinación entre áreas y la arquitectura para la toma de decisiones. En la base se encuentran las personas que generan los productos y/o servicios que la organización ofrece. Éstas son su núcleo operativo (Serrat, 2010). A menos que sean muy simples, las organizaciones suelen requerir de coordinadores que planeen y supervisen las actividades del núcleo operativo.

Mientras más compleja sea ésta, serán necesarios más coordinadores y parte del personal se organizará en subunidades especializadas que presten servicios al interior de la organización, al hacer, por ejemplo, tareas administrativas, dedicarse al manejo de la información o a la vinculación interinstitucional (Serrat, 2010). En ese sentido, un indicador del valor puesto en las evaluaciones es contar con una persona o área responsable de esa función.

Por otro lado, la estructura operacional de la organización influye bastante en la forma en que se toman decisiones en ella. Las organizaciones utilizan un estilo directivo y jerárquico, cuando sus líderes toman la mayor parte de las decisiones de forma unilateral. El estilo es adecuado cuando no se dispone de mucho tiempo y cuando las decisiones son sencillas. Sin embargo, es inapropiado cuando el líder necesita información del resto de los integrantes de la organización, o cuando la situación es compleja y de resultados inciertos (National Outdoor Leadership School, s.f.).

Dado que el estilo anterior tiende a disuadir el disenso y el intercambio franco de ideas (Kusters *et al.*, 2011), se considera que las organizaciones que aprenden basan su toma de decisiones en el trabajo grupal (Serrat, 2010). El estilo colaborativo parte de la siguiente idea: el conocimiento que se necesita para poder tomar decisiones complejas se desarrolla a través del intercambio de ideas entre individuos. Por ello es crucial que en la organización sea posible confrontar distintas posiciones de forma constructiva (Serrat, 2010).

Este estilo se fortalece cuando la organización tiene una estructura que permite la participación de sus integrantes en la toma de decisiones, sea al aportar información para que los líderes decidan o al decidir directamente a través de votaciones o consensos. La participación promueve que los integrantes de la organización valoren las evaluaciones, si saben que son un mecanismo que les permite influir en la toma de decisiones de la organización (Patton, 1997).

Por este motivo, de manera más amplia, se ha ligado el uso de evaluaciones al **liderazgo** de los que encabezan la jerarquía de comando. Después de todo, para que el aprendizaje se institucionalice en las organizaciones, este comportamiento necesita ser valorado por los líderes de la organización (Taut, 2007; Serrat, 2010; Preskill y Boyle, 2008; Kusters *et al.*, 2011; Taylor *et al.*, 2013; Cousins *et al.*, 2014). Esto es así, pues los líderes tienen el poder de decidir sobre las prioridades, la distribución de recursos y la estructura de incentivos. Por ello, a menos de que estos apoyen claramente una cultura de aprendizaje, será muy difícil establecerla. Puede existir un responsable de hacer o de delegar las evaluaciones, pero en última instancia son los directivos de una organización los encargados de alinear los valores e incentivos para que puedan usarse para aprender (David, s.f.).

Crear que el aprendizaje a partir de evaluaciones hará que la organización sea más efectiva, implica que los líderes están dispuestos a comprometer recursos y a permanecer involucrados incluso cuando las evaluaciones muestren que las cosas marchan mal (Kusters *et al.*, 2011). Para ello es necesario que los líderes diseñen y aprueben una estrategia que provea los medios y las oportunidades (Figura 5) para que el personal de la organización aprenda (Serrat, 2010).

Finalmente, de ellos depende, en gran medida, que se construya un aspecto crucial para que el aprendizaje a partir de evaluaciones pueda ocurrir: **relaciones de confianza**. A pesar de que su propósito sea el aprendizaje, las evaluaciones por definición sirven para hacer juicios de valor. Por esa razón, en una sociedad meritocrática, es dado esperar que generen ansiedad. El temor a una evaluación negativa parte de raíces psicológicas que se vinculan con la preocupación por el estatus, como la pérdida de credibilidad, respeto o aceptación; éstas son experiencias negativas que la mayor parte de las personas intentan evitar (Donaldson *et al.*, 2002).

La crítica genera ansiedad cuando, en un grupo, se utiliza la calidad de las ideas y del trabajo como un criterio para juzgar el valor de una persona. Si se tratan como un aspecto que permite definir a su propietario, es decir, como un medio para saber quiénes son las personas, entonces podemos entender por qué con frecuencia las personas interpretan que una amenaza contra sus ideas o su desempeño es una amenaza contra su persona. En consecuencia, defender las ideas propias se convierte en un tipo de defensa personal y admitir el error es cómo admitir que se es un fracaso (Reddy y Wilson, 2016).

Figura 5. Creando el motivo, la oportunidad y los medios para el aprendizaje.



Fuente: Modificado de Britton (2005) por Serrat (2010)

Por ello, para que los integrantes de una organización puedan involucrarse en la reflexión sobre lo que hacen, es necesario que existan altos niveles de confianza interpersonal, de forma que sea seguro para las personas el compartir abiertamente su opinión y experiencia (King, 1998; Serrat, 2010; Kusters *et al.*, 2011). Al contrario, cuando las personas sienten una excesiva preocupación por la crítica, dado que asumen que serán culpados por las fallas que se cometan, entonces se dificultará la reflexión pública sobre su práctica (Wild y Ramalingam, 2018). La confianza permite aprender de los errores, mientras que la cultura de culpar fomenta que éstos sean enterrados (Cuadro 16).

Cuadro 16. Respuestas organizacionales ante el error.

RESPUESTA DE AUTOENGAÑO	RESPUESTA DE DERROTA	RESPUESTA DE APRENDIZAJE
Definición del error		
Fallas en las decisiones estratégicas	Una fuerza que no se puede controlar	Una fuente de información.
Respuesta		
El error se oculta o alguien es culpado.	Se habla sobre el error de forma detallada, pero no se toma acción alguna	Se habla abiertamente del error y se intentan acciones correctivas.
Impacto en el liderazgo		
Los líderes son engañados (en parte) y piensa que todo va conforme al plan.	Los líderes no pueden promover cambios y la organización se inmoviliza.	Los líderes pueden basarse en lecciones aprendidas para tomar decisiones

Fuente: Korten y Klauss (1984) en Serrat (2010)

c) Sistemas de soporte

Los sistemas de soporte son el marco dentro del cual los integrantes de una organización laboran. Éstos definen quién hace qué y cómo ha de hacerlo. Por ende, hacen referencia a las actividades diarias, reglas, políticas y procedimientos que tiene que seguir el personal para desempeñar su trabajo. Se incluyen aquí los sistemas de información y comunicación, los cuales permiten recabar, guardar, analizar, compartir y usar información de forma sistemática. También el sistema de incentivos, que define lo que es recompensado y sancionado, así como el sistema de capacitación del personal, que facilita el desarrollo profesional de los integrantes de una organización.

Se presentan a continuación varios sistemas relacionados con la generación de evaluaciones.

-Sistema de información y comunicación

Este sistema indica los propósitos que la información tiene, de los cuales dependen los planes y procedimientos utilizados para manejarla. En ese sentido, si uno de los propósitos es demostrar transparencia financiera, con el objetivo de comprobar la implementación de los planes acordados, entonces se puede recurrir a reportes de resultados donde se estipulen los gastos que se efectuaron para realizar las actividades.

A su vez, si la intención es incrementar el entendimiento del contexto en el que se está interviniendo, entonces son útiles estudios diagnósticos para establecer líneas base o para comprender a profundidad las problemáticas.

Sin embargo, si el propósito es mejorar la práctica de la organización, entonces se requiere recabar, almacenar, analizar y compartir el conocimiento generado por sus integrantes. Esto permite preparar flujos de información basados en la pregunta sobre quién hace qué y por lo tanto quién necesita saber qué para desempeñarse mejor. La evaluación es la herramienta más reconocida para conseguirlo (Preskill y Boyle, 2008; Kusters *et al.*, 2011; Cousins *et al.*, 2014).

Algunos signos observables de que una cultura evaluativa se ha institucionalizado en procedimientos son:

- la presencia de un plan estratégico de evaluaciones que describa cómo, cuándo, por quién y en qué grado diferentes proyectos serán evaluados (Preskill y Boyle, 2008; Bourgeois y Cousins *et al.*, 2013).
- la planeación de espacios formales e informales para que el personal comparta conocimientos con diferentes áreas de la organización (Guijt, 2008; Ekboir *et al.*, 2009).
- la redacción de objetivos específicos, medibles, atribuibles, relevantes y ceñidos a un tiempo de cumplimiento (Mora y Antonie, 2012 en Stewart, 2014).
- la existencia de plataformas para facilitar el acceso, almacenamiento, transferencia y diseminación del aprendizaje (Serrat, 2010).

Sin embargo, si bien es importante que existan procedimientos de evaluación, hay que cuidar que no sean interpretados como reportes a llenar de forma rutinaria, pues esa actitud inhibiría su potencial como herramienta de aprendizaje (Raeside, 2011).

-Sistema de capacitación

Es necesario un trabajo intencional para crear y sostener procesos organizacionales que hagan que la producción y uso de evaluaciones de calidad sean una rutina en las organizaciones (Stockdill *et al.*, 2002: 8). Para ello, es conveniente una estrategia de capacitación que garantice la adquisición de habilidades y

herramientas esenciales, como el pensamiento crítico, el trabajo en equipo, la comunicación efectiva y el uso de métodos de recolección de información (Bourgeois y Cousins, 2013; Cousins *et al.*, 2014; Buckley *et al.*, 2015).

El desarrollo de capacidades en evaluación no es fácil, por lo que se requiere de un compromiso de largo plazo. Esto se opone a la implementación de cursos cortos, los cuáles predominan cuando el financiamiento proviene de un proyecto en específico (Lennie *et al.*, 2015). Este aspecto es especialmente importante si en la organización hay mucha rotación de personal. En esos casos la organización debería de proveer de oportunidades frecuentes de capacitación (Preskill y Boyle, 2008).

-Sistema de incentivos

Un sistema de incentivos define los comportamientos que se recompensan y que se sancionan al interior de una organización (Ekboir *et al.*, 2009). En esa línea, se considera valioso asegurar a los integrantes de una organización que serán reconocidos el tiempo y esfuerzo invertidos en la generación, socialización y uso de información proveniente de evaluaciones (Cousins *et al.*, 2008; Serrat, 2010; Cousins *et al.*, 2014). Si bien, normalmente se piensa en estímulos financieros como primas o incrementos de sueldo, hay un gran potencial en incentivos intangibles y beneficios no económicos, especialmente si se desea conseguir un involucramiento (Cuadro 17).

Cuadro 17. Incentivos intangibles y beneficios no económicos para el manejo basado en resultados.

Tipo de incentivo	Descripción
INTANGIBLE	<ul style="list-style-type: none"> • Reconocimiento personal • Reconocimiento público • Remoción de restricciones y delegación de autoridad para tener más flexibilidad
BENEFICIOS	<ul style="list-style-type: none"> • Traslado a congresos • Mejor espacio de oficina • Licencias y selección para capacitaciones • Horario de trabajo flexible

Fuente: Mayne, 2010.

-Recursos

Para que los sistemas que soportan el aprendizaje a partir de evaluaciones se puedan mantener, es necesario destinar recursos. En específico se requiere que los integrantes de la organización tengan suficiente tiempo para realizar las actividades

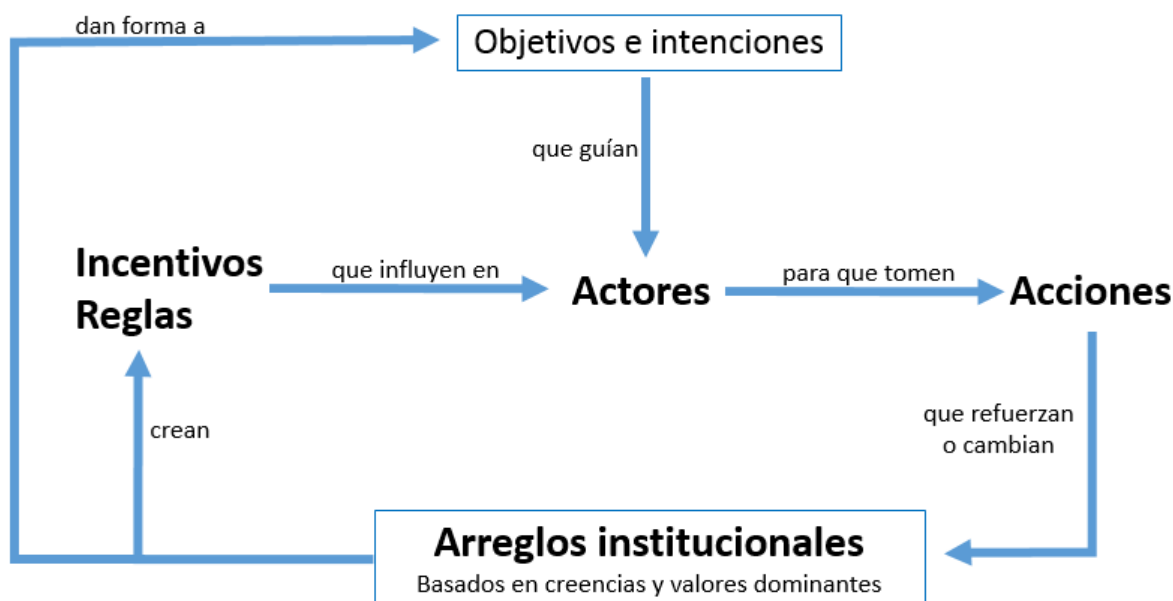
de evaluación, incluido el espacio para reflexionar, y un presupuesto definido, tanto para hacer evaluaciones como para implementar cambios (Cousins *et al.*, 2008; Guijt, 2010; Kusters *et al.*, 2011; Peersman *et al.*, 2015; Wild y Ramalingam, 2018).

1.2.2.3. Arreglos institucionales

La injerencia que tiene el contexto sobre la actuación de una organización ha sido estudiada por diversos investigadores del campo del desarrollo (Lewis *et al.*, 2003; Guijt, 2008; Eyben, 2008). Múltiples circunstancias que están fuera del control de una organización influyen sobre el tipo de decisiones que se pueden tomar en ella. Por ello, entender este entorno es crucial para el uso de las evaluaciones.

La mayor parte de las organizaciones operan bajo arreglos institucionales que norman su actuación (Figura 6). Los arreglos reflejan valores dominantes en una serie de mecanismos (p. ej. reglas, procedimientos o incentivos) que influyen en su comportamiento; es decir, las organizaciones ajustan sus procedimientos para adaptarse a su contexto (Ekboir *et al.* 2009). Cuando existe un desbalance en la determinación de los valores que orientan la acción, los arreglos reflejan las relaciones de poder entre distintas organizaciones (Lewis *et al.*, 2003).

Figura 6. Influencia de los arreglos institucionales sobre el comportamiento organizacional



Fuente: Modificado de Vermeulen *et al.* (2008).

Estos son algunos ejemplos relacionados con el objeto de la investigación:

- Un marco normativo nacional que establece la forma en que distintos órganos de gobierno han de llevar a cabo evaluaciones.
- Las reglas y procedimientos establecidos por un donante para evaluar el desempeño de sus donatarios.
- Los términos de referencia bajo los que una dependencia de gobierno solicita la evaluación de los proyectos, financiados por ésta, pero implementados por terceros.
- Los indicadores con los que una dependencia de gobierno evalúa el desempeño de instituciones académicas.

En todos los casos anteriores, las organizaciones encargadas de llevar a cabo proyectos no determinan las características y propósito de sus evaluaciones. Independientemente de si las organizaciones valoran las evaluaciones en sí, éstas son obligadas a adoptarlas por un mandato o por su dependencia hacia recursos externos. Por ello, el modelo del principal-agente (Ver 1.1.3.1) es útil para explorar las relaciones de poder entre las organizaciones. En primer lugar, para identificar al principal en la relación (quién o qué tiene el poder de llamar a cuentas). En segundo lugar, para describir los mecanismos prescritos, y finalmente para analizar el margen de autonomía institucional de los agentes. Después de todo, como señalaría Bourdieu (1990), el comportamiento de las organizaciones no está completamente atado a las estructuras; su lectura de las reglas del juego les permite decidir si cumplen, negocian o resisten las políticas que buscan gobernarlas.

Dentro de esta dinámica, los conceptos de isomorfismo coercitivo y mimético (DiMaggio y Powell, 1983) son valiosos para describir algunas circunstancias en las que se adoptan formas organizacionales externas. El isomorfismo coercitivo ocurre cuando las organizaciones son forzadas a adoptar formas debido a la dependencia de recursos. Mientras tanto, en el mimético, las organizaciones imitan formas que perciben les permitirán ser exitosas en un contexto dado, y por tanto, las adoptan de forma voluntaria.

Dicho lo anterior, es importante identificar los modelos hegemónicos de evaluación en el campo del desarrollo, así como los valores en los que se sustentan. Estos modelos establecen qué es lo que se ha de evaluar, cómo asignar valor a los proyectos, y cómo utilizar el conocimiento generado (Shadish, 1998 en Donaldson y Assam, 2015: 18). Por ello, si las organizaciones son inducidas a adoptarlos, es obvio que ejercen una influencia importante sobre su uso de evaluaciones.

En primera instancia, se describirán las características e influencia de evaluaciones sumativas que basan la rendición de cuentas en resultados. Sus efectos sobre el comportamiento de las organizaciones ha sido ampliamente documentado (Ebrahim, 2003; Eyben, 2015). Después, se describen las características de evaluaciones formativas que basan la rendición de cuentas en el aprendizaje.

a) Evaluaciones de desempeño y cambio proyectable

En la década de los setentas, surgió el enfoque de la Nueva Gestión Pública (NGP) como respuesta a las limitaciones del modelo de administración pública tradicional. El modelo burocrático asumía que el control jerárquico centralizado conducía a la eficiencia (González y Cerdán, 2008). Por ello, se confiaba en que, si las instrucciones se cumplían a cabalidad, entonces se conseguirían los objetivos. En ese sentido, los funcionarios rendían cuentas al demostrar que habían seguido los ordenamientos y entregado los productos esperados.

La NGP cuestionó la efectividad de ese modelo, y postuló que además de revisar que el dinero era utilizado de acuerdo a ciertas reglas, era necesario corroborar que se usaba de forma eficaz y eficiente (Eyben, 2015). Es decir, había que rendir cuentas con base en los resultados obtenidos, y no simplemente a partir del seguimiento de procedimientos. Los resultados se entienden, desde esta perspectiva, como el valor público generado por las intervenciones, al cubrir las necesidades e intereses de los ciudadanos (González y Cerdán, 2008).

Se han desarrollado distintos enfoques que comparten el principio de la obtención de resultados, como el Presupuesto basado en Resultados (PbR) o la Gestión para Resultados del Desarrollo (GpRD). Éstos aportan conceptos, procedimientos e instrumentos específicos para que las organizaciones generadoras de valor público sean más eficaces (González, 2013).

En ese sentido, estos enfoques buscan mejorar el funcionamiento de las organizaciones al dotarlas de prácticas gerenciales que fundamenten la gestión de los proyectos sobre decisiones racionales. Por ello, se ha dicho que la agenda de resultados ha contribuido a la profesionalización de muchas organizaciones con incipientes sistemas de monitoreo y evaluación (Segerholm, 2002 en Cousins *et al.*, 2015). Esto es así, pues la necesidad de rendir cuentas es un estímulo para tener discusiones sobre cómo medir el éxito, circunstancia que permite, a su vez, fortalecer la lógica de los proyectos y apreciar la importancia de la práctica evaluativa (Whitty, 2015).

En el campo de la conservación, la NGP ha enfatizado las evaluaciones de desempeño que miden el impacto con indicadores de calidad ecológica (Turnhout, 2009). Desde esta lógica, se ha propuesto que las convocatorias soliciten metodologías integrales de evaluación de impacto como requisito para el otorgamiento de recursos, de forma que se sustituya la tendencia de otorgar recursos para iniciativas que exageran *ex ante*, y sin evidencia alguna, sus posibles efectos (Baylis *et al.*, 2016).

Aun así, se ha señalado el predominio de evaluaciones simplificadas que sólo valoran la obtención de *outputs*, los cuales representan los niveles de logro esperados (Ebrahim, 2003). Si bien lo ideal sería que las organizaciones estimaran las evaluaciones como una herramienta que les permite conseguir sus objetivos, este tipo de evaluaciones se hacen sólo para satisfacer los intereses de los principales, siendo irrelevantes para la toma de decisiones al interior de las organizaciones (Cousins *et al.*, 2015; Ebrahim, 2003).

El supuesto que está detrás de este propósito de las evaluaciones, prevalente en el campo del desarrollo, es que el cambio es predecible y proyectable (Ebrahim, 2003; Reeler, 2007; Eyben, 2008). En ese sentido, este “enfoque de contrato” permite acordar por adelantado los resultados que se buscan, las estrategias que causarán el cambio, los recursos que se requerirán, y los medios a través de los cuales las partes del contrato sabrán si los resultados se han conseguido (Eyben, 2008). El rol de las evaluaciones es valorar el grado en el que se consiguieron los resultados previstos (Eyben, 2014).

Esta idea asume que los proyectos se implementan en contextos simples, caracterizados por la estabilidad y claridad de sus patrones de causa-efecto. Por ello, se percibe que el cambio es algo objetivo, cuantificable e independiente del entorno (Eyben, 2015). Para evaluar las intervenciones que operan en este tipo de contexto es suficiente la recolección rutinaria de datos sobre variables elegidas para compararlas con el desempeño proyectado, siendo sencillo contar con indicadores cuantitativos que permitan valorar el progreso acorde a los planes (Guijt, 2008).

Las campañas de vacunación son un ejemplo de intervenciones de este tipo, pues existe una alta predictibilidad entre la actividad (vacunar) y el efecto esperado (protección contra enfermedades). En esos casos, cumplir con las metas planteadas representa una aproximación suficientemente confiable del cambio que el proyecto ha producido.

Sin embargo, esta forma de proceder no funciona cuando se opera en contextos complejos, donde, debido a la gran cantidad de elementos interactuando entre sí de una manera no lineal, existe incertidumbre y desacuerdos sobre lo que hay que hacer (Eyben, 2008). En ellos, las relaciones entre causa y efecto son solo evidentes en retrospectiva y dependen en gran medida de las condiciones de partida. Este es el campo de lo no conocible, donde cada situación es única y el éxito pasado provee pistas insuficientes para definir la acción futura (Snowden y Boone 2007: 71).

La principal regla de los problemas complejos es que no pueden ser tratados como problemas simples (Poppendieck, 2002 en Guijt, 2008: 15). Aun así, en torno de proyectos de desarrollo, un problema recurrente es la simplificación de un contexto complejo para adoptar un estilo de manejo de comando y control. Algunos de los efectos nocivos que esta forma de hacer evaluaciones tiene sobre el comportamiento de las organizaciones son los siguientes:

-Dificultades para adaptarse

La lógica del cambio proyectable asume que la planeación racional de intervenciones y una ejecución correcta son condiciones suficientes para conseguir los efectos deseados (Eyben, 2008). Como consecuencia, las organizaciones enfrentan una serie de limitaciones para hacer adaptaciones sobre la marcha, como presupuestos con rígidos ritmos de ejecución o complicaciones administrativas para modificar cadenas de resultados (Wild y Ramalingam, 2018).

-Sesgo autorreferencial

Hacer que la complejidad aparezca como algo simple y aparentar, por ende, que el proyecto es infalible es una trampa para las organizaciones en el campo de desarrollo (Eyben, 2008). Con frecuencia, con el objetivo de asegurar el financiamiento, suelen presentarse propuestas románticas y apolíticas que simplifican la realidad y que prometen generar cambios sociales profundos (Guijt, 2008). Esto conduce a la implementación de proyectos que son poderosos discursivamente, pero que tienen pocas probabilidades de materializar los resultados que prometen (Mathews, 2011).

En este contexto, la presión por demostrar resultados puede generar un sesgo autorreferencial al utilizarse la evaluación para confirmar las convicciones previas y no para probar la verdad de las teorías de cambio que respaldan los proyectos. En ese sentido, como afirma Eyben (2008), la evaluación puede usarse tan sólo para verificar que se sigue en el camino y no para preguntar si se ha tomado el camino correcto.

El apremio por probar que el proyecto está funcionando “como debería” incentiva que se busque información que apoya los supuestos de origen, y que se omitan los datos que los contradicen (Mosse, 2004; Agudo, 2015). De este modo, la teoría del cambio de los proyectos y los impactos “esperados” se convierten en un punto de referencia para establecer lo que sucedió, produciendo datos distorsionados en lugar de valorar lo que realmente ocurrió con las personas involucradas (Batliwala y Pittman, 2010 en Bell y Aggleton, 2012). En otras palabras, el mapa se convierte en el mundo y se pasa de la política basada en evidencia a la “evidencia basada en la política” (Guijt, 2010: 279).

Esto no quiere decir que las desviaciones del camino pasen desapercibidas a los ojos de los implementadores y de sus evaluadores, sin embargo, esa información puede ser tratada como una verdad del rey desnudo, la cual es percibida por todo el mundo, pero se calla por ser políticamente incorrecta (Agudo, 2015: 204).

Después de todo, cuando los actores tienen poco control sobre los eventos, es mucho más práctico controlar la interpretación de lo que ocurrió, siendo las evaluaciones un excelente medio para dar un sello de aprobación a los modelos de intervención (Mosse, 2004; Mosse y Lewis, 2006; Agudo 2015).

-Sesgo de contabilidad

Un rasgo característico del control jerárquico de proyectos es la inclinación hacia resultados que son fáciles de medir y cuantificar en el corto plazo, pues esto facilita las evaluaciones basadas en estándares de desempeño (Edwards y Hulme 1996; Ebrahim 2003; Natsios, 2010; Eyben, 2014). Sin embargo, esta forma de evaluación podría afectar la orientación de los proyectos hacia la generación de cambios sociales de largo plazo. Esto es así, pues la preferencia por financiar proyectos con resultados medibles y sin problemas de implementación afecta el financiamiento de

proyectos que buscan generar cambios transformacionales²¹ (Eyben, 2014). Este objetivo suele ser un proceso social y político de larga duración, que con frecuencia falla y cuyos indicadores, menos tangibles y más ambiguos, son difíciles de medir.

En ese sentido, el sesgo de contabilidad implica, siguiendo a Natsios (2010), confundir el valor de un proyecto con su capacidad para ser medido desde estándares cuantitativos. Dicho de otra manera, este sesgo impide percibir que no todo lo que cuenta se puede contar.

Por otro lado, el sesgo de contabilidad genera “miopía normativa” en las organizaciones (Ebrahim, 2003) cuando éstas se enfocan tanto en cumplir con los resultados de corto plazo impuestos por la administración de un proyecto, que pierden de vista los objetivos mayores que se han trazado como organización. Esto es especialmente cierto cuando las metas, representadas con números, no dicen nada acerca de la calidad del trabajo realizado. Pueden haberse organizado 5 comités o reforestado 200 hectáreas, pero eso no quiere decir que los efectos esperados se conseguirán.

Este comportamiento se explica en un contexto donde las organizaciones son valoradas en función de los resultados que demuestran, y donde el buen desempeño organizacional se demuestra a través de “datos duros” que cuantifican el impacto. En cambio, “contar una historia matizada sobre el cambio social, que involucre dificultades contextuales, complejas alianzas y resultados intangibles, pero esenciales, no lleva a ningún lugar” (Guijt, 2010: 280).

Esta “lógica de mensurabilidad” ha sido señalada como una limitante de instrumentos de política climática como REDD+²² (Turnhout *et al.*, 2015). REDD+ utiliza la tonelada de carbono como una métrica estandarizada y cuantificable para realizar pagos basados en el desempeño. Al hacerlo, se ha puesto mucho énfasis en métodos de evaluación centrados en el carbono, a pesar de que, en teoría, la estrategia busca generar también beneficios sociales y de biodiversidad. Ante las dificultades para estandarizar la medición de este tipo de resultados, los cuales son muchas veces intangibles, es probable que exista la percepción de que el pago con base en resultados cuantificables es el único método objetivo y confiable para recompensar el desempeño (Turnhout *et al.*, 2015).

Sin embargo, el uso exclusivo de indicadores cuantitativos como medida del desempeño puede, en algunos casos, hacer que se surroge, al creer que el indicador es una mejor medida del desempeño real de lo que en realidad es. Surrogar es un fenómeno psicológico que expresa la tendencia con la que se pierden de vista los constructos estratégicos que las mediciones intentan representar, para actuar como si las medidas fueran los constructos de interés (Bentley, 2019).

²¹ La transformación institucional involucra intervenciones que, de forma deliberada, cuestionan el *status quo* para reformar las instituciones (Guijt, 2008: 15).

²² REDD+ se refiere a iniciativas que buscan reducir las emisiones provenientes de la deforestación y degradación forestal, conservar e incrementar los stocks de carbono en el bosque y gestionar los bosques de forma sostenible (FAO, s.f.).

Siguiendo con el ejemplo anterior, los sistemas de evaluación que se apoyan exclusivamente en la cuantificación del carbono no solo son un intento de representación del desempeño, también se convierten en un esquema de incentivos que dirige las decisiones de las organizaciones (Turnhout *et al.*, 2015). Si la rendición de cuentas se “carboniza”, entonces pueden soslayarse múltiples objetivos sociales y ambientales establecidos de forma retórica en la justificación de la pertinencia de REDD+.

-Comportamiento administrativo estratégico

En campos donde los resultados buscados son de largo plazo y donde la causalidad no es predecible, basar los incentivos en el cumplimiento de resultados cuantificables podría generar incentivos perversos²³ en el comportamiento de los agentes (Mayne, 2010). A través de diferentes adagios como la Ley de Campbell²⁴ o la Ley de Goodhart²⁵, se ha expresado la tendencia con la que los sistemas de rendición de cuentas, que miden y recompensan el desempeño, al hacerlo, promueven su distorsión. Esto es así, pues causan que aquellos que son valorados modifiquen su comportamiento para ajustarlo a las métricas utilizadas, con lo que alteran los resultados.

“La crítica es que los indicadores de desempeño, cuando se usan para el control, no son confiables pues no miden el desempeño en sí mismo. Distorsionan lo que se mide e influyen la práctica hacia lo que está siendo medido, menospreciado lo que no se mide.” (Jake Allen, en Eyben, 2013: 14)

Una de las trampas de la rendición de cuentas excesiva es que, al ser medidos de forma frecuente, los evaluados se hacen mejores en el cumplimiento de los requisitos, pero no necesariamente esto los hace desempeñarse mejor en el mundo real (Bovens *et al.*, 2008: 228).

Este comportamiento administrativo estratégico, al enfatizar la satisfacción de los requerimientos por sobre la necesidad de ser efectivos en el mundo real, erosiona la posibilidad de construir diálogos reflexivos entre los grupos involucrados (Cousins *et al.*, 2015). Eso explica por qué, en ocasiones

“las oficinas de evaluaciones y los operadores del programa suelen estar más interesados en que las evaluaciones arrojen datos positivos sobre el desempeño del programa, que en generar lecciones útiles para su mejora.” (Cejudo, 2011: 17)

²³ Un incentivo perverso es aquel que causa un resultado no intencionado, ni deseado, que va en contra de los intereses de los diseñadores del incentivo (Bentley, 2019).

²⁴ “Mientras más se use un indicador social cuantitativo para tomar decisiones, más sujeto estará a presiones de corrupción, y más apto será para distorsionar y corromper los procesos sociales que intentaba monitorear” (Campbell, 1976: 34).

²⁵ Llamada así tras el economista Charles Goodhart. El adagio se ha sintetizado de la siguiente manera: “Cuando una medida se convierte en un objetivo, deja de ser una buena medida” (Strathern, 1997).

Este tipo de comportamiento no necesariamente quiere decir que las organizaciones conspiran cínicamente contra sus propios objetivos. Algunas de ellas, comprometidas con el cambio y no con la simulación, pero presionadas por un esquema de rendición de cuentas a comportarse contra su buen juicio, pueden cumplir con lo solicitado para asegurar el financiamiento, mientras hacen en la medida de lo posible lo que les parece más apropiado (Eyben, 2014).

En este escenario, sea por temor o con la finalidad de sostener beneficios, las organizaciones pueden seguir procedimientos impuestos, como las evaluaciones, sin internalizar su lógica o los valores que les subyacen (Eyben, 2008). Esto hace que herramientas que bajo otras condiciones serían potencialmente útiles se conviertan en un instrumento coercitivo que reducen el espacio de elección y la innovación (Eyben, 2015).

-El miedo a fallar

Diversos autores han afirmado que el miedo a la evaluación es una de las principales barreras que impiden el uso de la información que proviene de ellas (Behn, 2002; Donaldson *et al.*, 2002). Es posible explicar este comportamiento, dado que normalmente se temen las valoraciones negativas²⁶, por lo que son eventos que se desea evitar. Es por ello que las evaluaciones que tienen por propósito controlar o valorar el desempeño suelen ser vistas como una amenaza.

La creencia de que es posible predecir que el cambio sucederá en la dirección deseada ha favorecido evaluaciones enraizadas en el temor de no conseguir los objetivos preestablecidos (Guijt, 2008). El resultado es un sesgo hacia lo que es de corto plazo y menos riesgoso, así como un fuerte desincentivo para actividades innovadoras, que, si bien más propensas a las fallas, podrían encontrar mejores estrategias para conseguir los objetivos (Perrin, 2002; Bovens *et al.*, 2008).

De esta manera, la tensión entre cumplimiento y aprendizaje dificulta que se logre un balance entre ambos propósitos, dado que, aun cuando la organización haya aprendido cómo hacer mejor las cosas, en caso de fallar las consecuencias negativas son muy altas (por temor a la pérdida de credibilidad o financiamiento).

Behn expresa esta situación de la siguiente manera:

“Cuándo los empleados públicos piensan en las consecuencias de su trabajo, están más preocupados por evitar errores que con certeza producirán una sanción, que por conseguir logros que podrían generarles alguna recompensa” (2002: 12).

En este sentido, se ha documentado el problema, común a ONG´s y dependencias de gobierno, de “fallar en el aprendizaje basado en fallas”, causado por los fuertes incentivos por mostrar resultados y por encerrar y olvidar las experiencias negativas (Forss *et al.*, 2002).

²⁶ Relacionadas con experiencias desagradables, como la pérdida de aceptación, la vergüenza, el ridículo o el rechazo (Donaldson *et al.*, 2002).

b) Evaluaciones formativas y manejo adaptativo

Las distorsiones que se han descrito dejan ver que, en ciertas situaciones, más rendición de cuentas no conduce a una mayor efectividad (Bovens *et al.*, 2008). Sin embargo, esto no implica que hay que dejar de procurar la rendición de cuentas basada en resultados, el *quid* está en cómo hacerlo.

Es justo que en el campo del desarrollo las organizaciones rindan cuentas a través de su efectividad, pero para dar resultados en entornos complejos se requiere espacio para la experimentación y la adaptación. Por ello, el manejo adaptativo debería considerarse un mecanismo de rendición de cuentas, pues los ajustes buscan incrementar el valor generado a cambio del dinero recibido (Wild y Ramalingam, 2018).

Para ampliar la influencia de este tipo de rendición de cuentas basada en el aprendizaje, se requiere identificar su área de aplicación. Después de todo, no todo es complejo, ni impredecible, por lo que algunos procesos pueden evaluarse con base en cumplimiento de estándares prestablecidas, mientras que otros pueden evaluarse al reunir personas para que juntas analicen una situación confusa, con la finalidad de adaptarse y poder responder mejor (Guijt, 2010).

Lamentablemente, esta visión va en contra de las tendencias que dominan los arreglos institucionales del campo del desarrollo. Las exigencias del financiamiento basado en proyectos hacen que internacionalmente exista un desbalance que favorece el uso de evaluaciones con el propósito de controlar la administración de los proyectos (Cousins *et al.*, 2015).

Aun así, un número creciente de fundaciones se están convirtiendo en aliadas de sus donatarios, pues están transitando de un modelo de rendición de cuentas basado en la exigencia de resultados de corto plazo, a uno en el que se practica la empatía y se construyen relaciones en torno de la importancia de aprender juntos para incrementar el impacto (Darling, 2012).

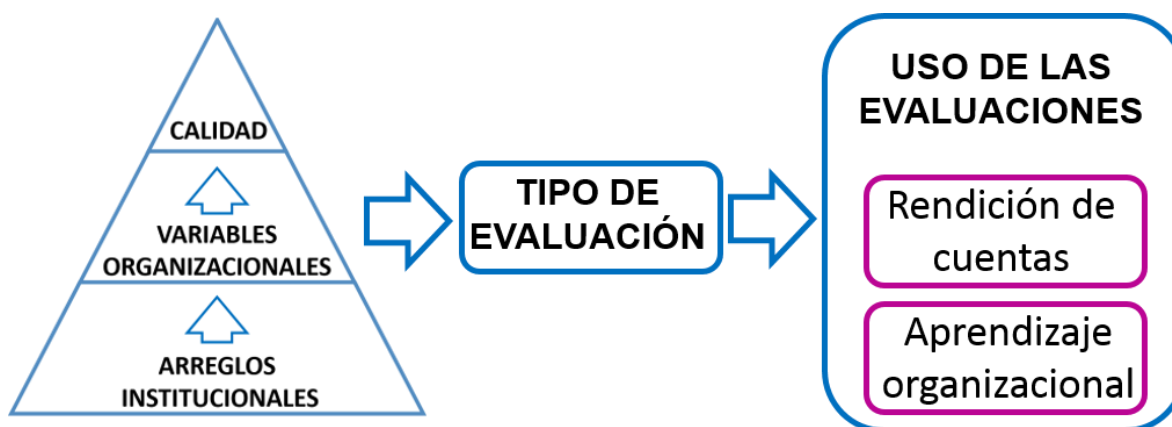
En ese sentido, es evidente el rol central que tienen los financiadores y reguladores en la promoción del aprendizaje a partir de evaluaciones. Por un lado, podrían promover esquemas de rendición de cuentas hacia abajo al garantizar el involucramiento de los beneficiarios, de forma que participen en la valoración de los resultados y la co-creación de soluciones (Ebrahim, 2003). Por el otro, podrían promover que las organizaciones, orientadas hacia los cambios que desean intencionar, planeen como conseguirlos y evalúen la efectividad de sus acciones para ajustarlas y desempeñarse mejor (Mayne, 2010).

Manifestaciones concretas de lo anterior son el financiamiento de proyectos multi-
anuales, la flexibilidad en el gasto a través de categorías presupuestales amplias, la existencia de fondos especiales para responder a eventos inesperados o financiar ajustes, el uso de evaluaciones participativas y el uso de evaluaciones intermedias que reportan fallas y alimentan ciclos regulares de reflexión (Wild y Ramalingam, 2018).

1.3 SÍNTESIS DE CONCEPTOS CLAVE

En este capítulo se presentó una revisión amplia de conceptos e ideas relativos a la evaluación y el uso de evaluaciones. A continuación, se utiliza un diagrama para sintetizar los conceptos clave que se emplean para responder las preguntas de investigación (Figura 7).

Figura 7. Vinculación causal de los conceptos clave empleados en la investigación



Fuente: Elaboración propia

La figura representa la vinculación causal de los conceptos centrales de la tesis. En el extremo izquierdo, la pirámide unifica el conjunto jerárquico de factores que, se asume, tienen incidencia sobre:

- a) los tipos de evaluación realizados por las organizaciones y, en última instancia,
- b) sobre la forma en que éstas son utilizadas.

A continuación, se enlistan los conceptos que componen cada uno de los bloques del diagrama, refiriendo al lector al número de página en donde se explican con detalle.

- Factores que inciden en la producción y uso de evaluaciones (Cuadro 18)
- Tipos de evaluación (Cuadro 19)
- Usos de la evaluación (Cuadro 20)

Cuadro 18. Factores que inciden en la producción y uso de evaluaciones.

NIVEL	CONCEPTO	PREGUNTAS CLAVE
Calidad de las evaluaciones	<ul style="list-style-type: none"> • Diseño de proyectos evaluables (p. 43) • Desarrollo de evaluaciones (p. 44) • Uso del proceso (p. 45) 	<p>¿Las organizaciones cuentan con representaciones de la cadena de resultados esperada con los proyectos?</p> <p>¿Hay coherencia entre el propósito de la evaluación y las metodologías empleadas?</p> <p>¿Las evaluaciones promueven la participación del personal de la organización que implementa los proyectos en el análisis y uso de los resultados?</p>
Variables organizacionales	<ul style="list-style-type: none"> • Responsables de evaluación en la estructura operativa (p. 51) • Sistema de información y comunicación (p. 52) • Sistema de capacitación (p. 52) • Recursos (p. 53) 	<p>¿La organización cuenta con personal especializado en la realización de evaluaciones?</p> <p>¿Cuál es la disponibilidad de recursos para realizar evaluaciones?</p> <p>¿Existe un plan estratégico de evaluaciones en la organización?</p> <p>¿Existen espacios para que el personal intercambie conocimientos?</p> <p>¿Existen plataformas para facilitar el almacenamiento e intercambio de información proveniente de evaluaciones?</p> <p>¿Se ha capacitado al personal en la realización de evaluaciones?</p>
Arreglos institucionales	<ul style="list-style-type: none"> • Influencia del entorno de la organización en la determinación del propósito y alcance de las evaluaciones (p. 54) • Signos observables de evaluaciones sumativas en relaciones de rendición de cuentas (p. 56-61) <ul style="list-style-type: none"> - dificultades para adaptarse, - sesgo autorreferencial, - sesgo de contabilidad, - comportamiento administrativo 	<p>¿Cuál es el margen de maniobra para determinar el propósito y alcance de las evaluaciones?</p> <p>¿Es posible modificar sobre la marcha la cadena de resultados y el ejercicio presupuestal?</p> <p>¿Es posible informar resultados negativos?</p> <p>¿Se temen las valoraciones negativas?</p>

	<p>estratégico</p> <ul style="list-style-type: none"> • Signos observables de evaluaciones formativas en relaciones de rendición de cuentas (p. 62) 	<p>¿Se cuenta con proyectos multi- anuales que usan evaluaciones intermedias que reportan fallas?</p> <p>¿El énfasis se pone en variables cuantificables?</p>
--	--	---

Fuente: Elaboración propia

Cuadro 19. Tipos de evaluación.

CONCEPTO	PREGUNTAS CLAVE
<ul style="list-style-type: none"> • Evaluación <i>ex ante</i> (p. 23) • Evaluación de desempeño (p. 27) • Evaluación de impacto (p. 27) • Evaluación de la efectividad organizacional (p. 32) 	<p>¿Cuáles son las características de las evaluaciones realizadas?</p>

Fuente: Elaboración propia

Cuadro 20. Usos de la evaluación.

CONCEPTO	PREGUNTAS CLAVE
<ul style="list-style-type: none"> • Sumativo (p. 25-26) - Rendición de cuentas con base en la entrega de productos esperados - Rendición de cuentas con base en resultados • Formativo (p. 28-30) - Aprendizaje de un ciclo - Aprendizaje de dos ciclos • Uso simbólico (p. 39) 	<p>¿Quién es el usuario principal de las evaluaciones y para qué utiliza la información?</p> <p>¿Es la evaluación un producto final?</p> <p>¿Es la evaluación parte del ciclo de manejo de los proyectos?</p> <p>¿Se utilizan las evaluaciones con fines simbólicos, como incrementar la credibilidad para la gestión financiera?</p> <p>¿Más allá de la información contenida en ella, participar en la evaluación permitió a los involucrados aprender?</p>

Fuente: Elaboración propia

Capítulo dos

Abordaje metodológico para estudiar el uso de evaluaciones

Abstract cleverness only separates the thinker from the world of reality, and that world is in a desperate condition now because of too many who think too much and care too little.

-Benjamin Hoff-

Este capítulo explica la metodología utilizada para dar respuesta a las preguntas de la investigación. En un primer apartado se describen las estrategias que se han empleado para estudiar el uso de evaluaciones en organizaciones del campo del desarrollo. Ese marco permite explicar las decisiones que se tomaron al diseñar la investigación. En un segundo apartado se describe cada uno de los componentes de la investigación, presentada como una serie de pasos en un diagrama de flujo. El capítulo finaliza con un análisis de las fortalezas y limitaciones de la metodología seleccionada.

2.1 ESTRATEGIAS PARA ESTUDIAR EL USO DE EVALUACIONES

Como objeto de estudio, el uso es el área más explorada en el campo de las evaluaciones (Christie, 2007). Se cuenta con investigaciones en torno de múltiples áreas, incluyendo la conservación, el desarrollo y el extensionismo agrícola (Guijt, 2008; Lamm *et al.*, 2011; Jacobson *et al.*, 2011). Además, el fenómeno se ha estudiado en muy diversos contextos organizacionales, desde fundaciones y dependencias de gobierno con grandes estructuras operativas y múltiples programas, hasta pequeñas organizaciones no gubernamentales enfocadas en la implementación de proyectos (Alkin y King, 2016).

A pesar de la recurrencia con la que el tema se ha estudiado desde hace cincuenta años, es muy complejo generalizar las características que más se relacionan con el uso de evaluaciones (Weiss, 1998; Johnson *et al.*, 2009). Al intentar entender por qué una organización hace lo que hace, los investigadores enfrentan el traslape de una serie de variables técnicas, psicológicas y políticas que aún no son bien comprendidas. Por ejemplo, a nivel institucional, existen avances modestos en el estudio de la influencia del entorno normativo sobre la forma en que las organizaciones adoptan ciertas prácticas (Camfield *et al.*, 2014). Mientras tanto, en la escala organizacional, hace falta entender mejor los efectos que tienen aspectos específicos de la cultura organizacional sobre el desempeño de las organizaciones (Asian Development Bank, 2008).

Debido a la influencia del contexto, son recurrentes los estudios de caso que vinculan datos empíricos con proposiciones teóricas (Christie y Vo, 2015). Esto permite contrastar las circunstancias de escenarios distintos, con la finalidad de identificar regularidades y diferencias. De ahí que abundan estudios que compilan factores asociados al uso (Leviton y Hughes, 1981; Cousins y Leithwood, 1986; Johnson *et al.*, 2009; Cardozo, 2015).

A pesar de esos esfuerzos de integración, hay que tener cuidado con las generalizaciones, pues la diversidad de los métodos empleados en los estudios dificulta la agregación de los hallazgos (Brandon y Singh, 2009; Cousins *et al.*, 2004). Por ende, la acumulación de conocimiento en este campo se orienta a conformar “teorías de patrones” (Lincoln y Guba, 2004) que marcan tendencias y permiten proponer principios generales (Alkin y King, 2019).

Independientemente, buena parte del valor del conocimiento generado se encuentra en su validez interna, dada la frecuencia con la que este tipo de investigación intenta incidir en la mejora de las organizaciones. En estos casos, el mérito de una investigación no se encuentra en la generación de conocimiento extendible a circunstancias externas, sino en su contribución para resolver los problemas específicos que estudia (Guba y Lincoln, 2002).

Dicho lo anterior, es posible agrupar los estudios clásicos del uso en dos líneas generales de investigación: a) el estudio de la influencia de las evaluaciones y b) el estudio de las organizaciones como usuarias de evaluaciones. En la primera de ellas, los estudios se han enfocado en el vínculo entre recomendaciones de política pública contenidas en evaluaciones externas y las decisiones de funcionarios de gobierno de alto nivel. Su objetivo es hacer un análisis *ex post* de los efectos ocasionados y a partir de ahí hacer correlaciones con respecto de las características de la evaluación, el evaluador y el usuario de las evaluaciones (Pérez, 2015). En México, la mayor parte de las investigaciones se concentran en esta línea, siendo lideradas de forma notable por el Centro para el Aprendizaje en Evaluación y Resultados para América Latina y el Caribe²⁷ (Cejudo y Maldonado, 2011; Pérez, 2015; Cejudo y Abarca, 2016).

Para estudiar la influencia se han empleado encuestas autoadministradas, entrevistas semi-estructuradas a informantes clave y el análisis documental (González, 2016; Pérez, 2015; Cardozo, 2015). Las primeras dos técnicas sirven para recabar la opinión de los usuarios y la última permite reconocer si los cambios en el diseño de los programas tienen relación con las recomendaciones vertidas en las evaluaciones. Una vez sistematizada la información, se pueden realizar inferencias estadísticas para identificar la influencia de diferentes variables (Pérez, 2015; Cardozo, 2015).

Cuando la información existe y se tiene acceso a ella, el análisis documental es la técnica más propicia para evidenciar la influencia de evaluaciones (Weiss, 1998). Sin embargo, los textos en sí ofrecen pocos detalles sobre cómo funcionan las organizaciones en la práctica. Por ello, si el estudio se restringe a documentos las organizaciones permanecen como “cajas negras” en donde los diversos intereses e incentivos de sus integrantes quedan poco descritos (Lewis *et al.*, 2003: 546)²⁸.

Mientras tanto, los estudios centrados en las organizaciones consideran que las evaluaciones son parte del sistema de aprendizaje de una organización. Por ello, las investigaciones exploran los factores asociados con la capacidad de las organizaciones para hacer y usar evaluaciones (Cousins *et al.*, 2004; Taut, 2007; Taylor *et al.* 2013; Cousins *et al.*, 2014). En este sentido, las metodologías se enfocan en medir variables ligadas con la capacidad para aprender a partir de

²⁷ A nivel global, la iniciativa CLEAR (Centers for Learning on Evaluation and Results) es una estrategia encabezada por el Grupo Independiente de Evaluación del Banco Mundial. Su objetivo es fortalecer las capacidades de Monitoreo y Evaluación para mejorar el desempeño de las políticas públicas en países en vías de desarrollo.

²⁸ Aun así, dada la dificultad para acceder a la vida cotidiana de las organizaciones, muchas investigaciones no tienen otra alternativa (Eyben, 2009: 77).

evaluaciones (Preskill y Torres, 2000; Cousins *et al.*, 2008; Taylor *et al.*, 2013; Cousins *et al.*, 2014).

Para estudiar el fenómeno se han empleado entrevistas-semiestructuradas con actores clave y encuestas autoadministradas (Botcheva *et al.*, 2002; Cousins *et al.*, 2008; Bourgeois y Cousins, 2013). Sin embargo, una de las limitaciones de las encuestas es la posibilidad del sesgo de interés personal. Si no se requiere evidencia para soportar lo afirmado, entonces los encuestados pueden exagerar la información reportada (Poister y Streib, 1999). Por ende, la información obtenida corre el riesgo de tener poca relación con lo que sucede en realidad. Por otro lado, si la encuesta no se aplica a diferentes integrantes de la organización, es posible perder la perspectiva de parte del personal, dado que la percepción sobre las evaluaciones puede diferir al interior de la organización (Taylor *et al.*, 2013).

También se han empleado estudios de caso reflexivos y técnicas de investigación-acción, especialmente para analizar los factores involucrados en el desarrollo de capacidades en evaluación (Taut, 2007; Christie y Vo, 2015). El objetivo es obtener conocimientos sobre cómo facilitar procesos de aprendizaje a partir de evaluaciones en las organizaciones (Preskill y Boyle, 2008).

Si bien estas dos líneas de investigación han aportado conocimientos sobre cómo a) los evaluadores, b) la ejecución de evaluaciones y c) distintas variables organizacionales tienen un rol importante en el uso de evaluaciones, en sus análisis se describe muy poco la influencia del entorno institucional dentro del que las organizaciones operan. Por ello, se considera que el estudio sobre el uso de evaluaciones podría complementarse con metodologías provenientes de estudios críticos del desarrollo (Cuadro 21).

Cuadro 21. Diferentes formas de estudiar el uso de evaluaciones en el campo del desarrollo.

LÍNEA DE INVESTIGACIÓN	OBJETO DE ESTUDIO	METODOLOGÍA
Estudio de la influencia de evaluaciones	Uso de evaluaciones externas en la toma de decisiones de funcionarios de gobierno	Análisis documental Encuestas autoadministradas Entrevistas-semiestructuradas
Estudios centrados en las organizaciones como usuarias de evaluaciones	Capacidad de las organizaciones para hacer y usar evaluaciones	Encuestas autoadministradas Investigación-acción
Estudios críticos del desarrollo	Influencia del entorno institucional	Técnicas de investigación etnográfica

Fuente: Elaboración propia

En este tipo de investigaciones, el desarrollo es entendido como un discurso que ordena y regula lo que puede y no puede hacerse (Mosse y Lewis, 2006; Lewis *et al.*, 2003). Por ello, al estudiar evaluaciones, estas no se entienden simplemente como una herramienta para mejorar la toma de decisiones, sino como parte de un entramado de relaciones de poder que influye sobre el trabajo que desempeñan los involucrados en organizaciones de desarrollo (Mosse y Lewis, 2006: 3). Algunos de los aspectos analizados han sido la dinámica de relaciones de rendición de cuentas y la influencia que ejerce la agenda de resultados sobre el comportamiento de las organizaciones (Ebrahim, 2003; Mosse, 2004; Eyben, 2008; Guijt, 2008; Okail, 2009; Agudo, 2015).

Para estudiar el proceso se han mezclado diferentes técnicas de investigación etnográfica²⁹ como la observación participante (Mosse, 2004; Okail, 2009; Mathews 2011; Agudo, 2015), investigación participativa (Eyben *et al.*, 2007; Guijt, 2008) y entrevistas semi-estructuradas (Guijt, 2008; Mathews 2011). Estas técnicas permiten profundizar el análisis al explorar cómo y por qué el comportamiento y la acción son moldeados por reglas, procedimientos, valores y relaciones de poder (Lewis, 2001; Mosse, 2004).

Sin embargo, el uso de métodos etnográficos ha levantado cuestionamientos éticos en el estudio de organizaciones en el campo del desarrollo. Esto se ha debido, principalmente, porque el investigador mantiene una “agenda oculta” al no develar sus pretensiones a las personas con las que interactúa (Mosse, 2005). En este contexto, se considera que es responsabilidad del científico social tener el consentimiento de los informantes, y más aún en los casos en los que la información podría dañar su reputación (Devi, 2005). La posibilidad de romper la confianza es una implicación a anticipar y evitar, especialmente cuando el investigador planea mantener una relación de trabajo con los informantes al finalizar la investigación.

Otra crítica hecha a los estudios etnográficos ha sido su falta de vinculación constructiva, dado que varios investigadores consideran válido producir ideas sobre el proceso de desarrollo sin un objeto normativo, es decir, sin buscar que conduzcan al diseño de mejores intervenciones (Mosse, 2004; Mosse y Lewis, 2006). Sin embargo, esta postura ha sido rebatida desde la propia etnografía, a través de enfoques como la **Perspectiva de Aprendizaje** (PdA), la cual propone la existencia de un área interesante de uso práctico si se combinan la crítica de la etnografía con la instrumentalidad de aquellos que desean mejorar la práctica de las organizaciones (Eyben, 2009: 84)

En ese sentido, su propósito es facilitar la reflexión³⁰ conjunta del investigador y los actores implicados al tematizar situaciones problemáticas. La finalidad es examinar el comportamiento y explorar vías para modificar estructuras institucionales que bloquean objetivos que se valoran. Por ello, la perspectiva se encuentra en el campo

²⁹ La etnografía agrupa diferentes métodos que se basan en un contacto directo y sostenido con otros actores, generando como resultado descripciones detalladas de los encuentros (Willis y Trondman, 2002).

³⁰ La reflexividad se refiere a la capacidad de un individuo para analizar un proceso social dentro del que está envuelto, pero que antes le pasaba desapercibido. Ese reconocimiento lo transforma y le permite explorar sus posibilidades de cambiar el proceso social que analiza.

de interacción entre agencia y estructura al asumir que, aunque moldeados por el contexto social, los integrantes de organizaciones tienen la autonomía suficiente para decidir si contribuyen en la reproducción de estructuras existentes o si juegan una parte en cambiar la situación para mejor.

Una limitante de este tipo de investigación es que requiere un ambiente organizacional hospitalario. En su ausencia, el esfuerzo del investigador para invitar a otros a ser tan reflexivos como él puede desincentivar el proceso completo de descubrimiento (Eyben, 2009: 90). No es sencillo que organizaciones de desarrollo estén interesadas en participar en un proceso reflexivo sobre su propia actuación, especialmente si se analizan inconsistencias con respecto de la imagen que desean proyectar al exterior. En la balanza entra la percepción de las posibilidades de promover un cambio³¹, las consecuencias negativas que podría acarrear y los beneficios a obtener. Por ello, es necesario estar atentos a las coyunturas que posibilitan este tipo de intervenciones.

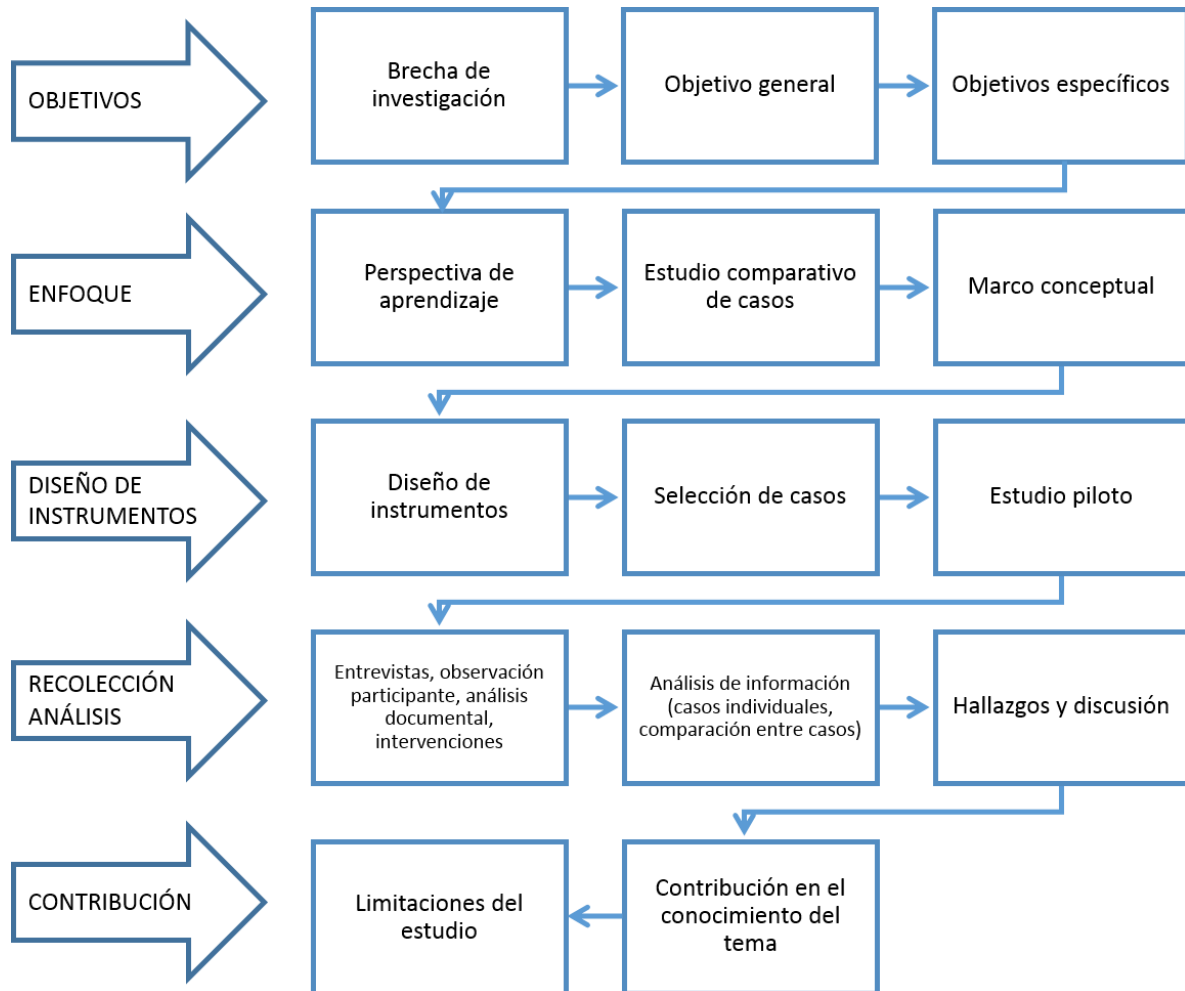
2.2 DISEÑO DE LA INVESTIGACIÓN.

En este apartado se presenta la secuencia lógica de la investigación (Figura 8). En primer lugar, se describe la brecha de investigación detectada, y a partir de ella se presentan los objetivos del proyecto. Después, se expondrán las razones por las que se decidió adoptar, cuando fuera posible, el enfoque de la PdA en un estudio comparativo de casos. Ligado a lo anterior, se desarrolla el marco conceptual que concentra las variables a estudiar.

En un momento posterior, se detalla la fase de diseño de instrumentos, incluyendo los criterios para seleccionar los casos, la estrategia de acercamiento a las organizaciones y los ajustes hechos a partir de una prueba piloto. Finalmente, se detalla la forma y magnitud en que se emplearon las técnicas de recolección de información y se analizan las limitaciones de la metodología elegida.

³¹ Después de todo, sin importar lo reflexivo que se sea, el poder y la posición en el campo pueden impedir el impulso de los cambios (Eyben, 2014).

Figura 8. Secuencia lógica de la investigación.



Fuente: Elaboración propia

2.2.1 Brechas de investigación

La tesis atiende cuatro **brechas en la comprensión del fenómeno** que justifican llevar a cabo la investigación. Estas se identificaron al recuperar el conocimiento acumulado en torno del objeto de estudio. Si bien ya fueron presentadas en la introducción, se repiten a continuación:

1. A nivel internacional, se han hecho llamados para comprender mejor cómo se manifiesta en la práctica el manejo adaptativo en distintas organizaciones. A su vez, se ha solicitado incrementar la base de evidencia sobre las condiciones necesarias para institucionalizarlo, de forma que logre sus objetivos de

iteración programática y toma de decisiones basadas en evidencia. Por ello, se requieren más estudios de caso que ilustren el fenómeno (Guijt, 2008; Wild y Ramalingam, 2018; Pasanen y Barnett, 2019).

2. En México, no se ha estudiado la forma en que las universidades públicas evalúan sus proyectos de investigación aplicada. El énfasis se ha puesto, en cambio, en el análisis de los efectos del sistema de evaluación de la política científica federal sobre el comportamiento de los investigadores (Galaz y Gil, 2013; Suárez y Muñoz, 2004).
3. No hay investigaciones sobre el sistema de evaluación diseñado por la Comisión Nacional de Áreas Naturales Protegidas (CONANP), dependencia federal responsable del manejo de ANP en México. Dado que normalmente se considera que los usuarios de las evaluaciones son funcionarios de gobierno de primer nivel (Cardozo, 2015), las investigaciones se han centrado en esa escala, al estudiar las evaluaciones externas de sus programas de subsidio (Orta *et al.*, 2013; Sotelo *et al.*, 2015).
4. En México, el estudio sobre el uso de evaluaciones se ha concentrado en la política de evaluación del gobierno federal (Cejudo y Maldonado, 2011; Cardozo, 2015). Sin embargo, existen pocas investigaciones sobre el uso de evaluaciones en la escala local³², con organizaciones directamente involucradas en la implementación de proyectos de conservación y desarrollo. Es importante estudiar el fenómeno a ese nivel, pues los integrantes que operan las intervenciones deberían ser los principales usuarios de evaluaciones para manejar a través de adaptaciones.

2.2.2 Objetivos

General

- Comparar la forma en que distintas organizaciones hacen y usan evaluaciones en una reserva de la biosfera del centro de México.

Específicos

- Identificar los factores que promueven e inhiben el uso de evaluaciones para dirigir la toma de decisiones de los guardaparques en una reserva de la biosfera del centro de México.

³² Las investigaciones existentes abordan los sistemas de evaluación de entidades federativas y de algunos municipios capitales (Maldonado y Galíndez, 2013; Consejo Nacional de Evaluación de la Política de Desarrollo Social, 2016).

- Identificar los factores vinculados a la producción y uso de evaluaciones en un centro público de investigación agrícola.
- Realizar un análisis comparativo de la forma en que distintos tipos de organización hacen y usan evaluaciones en una reserva de la biosfera del centro de México.
- Establecer el grado en que las organizaciones han adoptado evaluaciones como una herramienta de aprendizaje organizacional.
- Hacer recomendaciones específicas para fortalecer la capacidad evaluativa de cada organización estudiada.

2.2.3 Perspectiva de aprendizaje

Para estudiar el fenómeno, se decidió adoptar el enfoque de la PdA (Eyben, 2009). Estas son las razones:

1. El investigador, independientemente del proyecto, ha tenido y planea tener una relación de trabajo con el personal del resto de las organizaciones. En ese contexto, se consideró valioso combinar el análisis del fenómeno con la intención de mejorar la práctica de las organizaciones.
2. Se concibe que los científicos, más allá de ser productores de conocimientos, deben ser actores comprometidos con el uso del conocimiento para transformar situaciones que consideran deben cambiar (Guba y Lincoln, 2002).

En consecuencia, en el marco de un estudio comparativo de casos, se estudió el fenómeno al combinar la revisión de documentos con técnicas etnográficas, y, en los casos en los que el investigador identificó coyunturas para intervenir, se facilitó un proceso reflexivo con personal de las organizaciones y se generaron propuestas concretas para mejorar su sistema de evaluación.

2.2.4 Estudio comparativo de casos

El proyecto se basó en el método de estudios de caso para investigar los factores relacionados con la producción y uso de evaluaciones en diferentes organizaciones que implementan proyectos de manejo de recursos naturales en la Reserva.

Los estudios de caso son un tipo de investigación cualitativa que analiza fenómenos en su propio contexto, lo que permite relacionar datos empíricos con planteamientos teóricos (Yin, 2009). Por ello, se considera que los casos son útiles para clarificar, redefinir o ampliar los supuestos provenientes de la literatura sobre el uso de evaluaciones (Stake, 1995; Guijt, 2008). Para describir las características de cada caso, y para hacer comparaciones entre ellos, es necesario seleccionar un conjunto de variables que, en teoría, se relacionan con la producción y uso de las evaluaciones. En ese sentido, a continuación se presenta el marco conceptual de la investigación.

2.2.5 Marco conceptual

Un marco conceptual, más que una colección de conceptos, es una herramienta que permite organizar la forma de abordar un objeto de estudio. En ese sentido, provee una estructura lógica para considerar y ponderar componentes que se presume tienen relación con el objeto de estudio, presumiendo una relación entre ellos (Jabareen, 2009).

A partir de una revisión de literatura, se elaboró un marco conceptual que reúne las categorías seleccionadas para estudiar el fenómeno (Ver Figura 7, Cuadro 18, 19 y 20). Este representa la influencia de una serie de factores representados de forma jerárquica como causa del tipo de evaluación y uso que realizan las organizaciones.

La jerarquización de factores por niveles de influencia no se ha propuesto como marco conceptual en investigaciones similares, las cuáles se enfocan solo en variables relacionadas con las evaluaciones, el evaluador o las organizaciones (Cousins *et al.*, 2008; Cousins *et al.*, 2014) pero sin integrar un análisis de la influencia de los arreglos institucionales que enmarcan la actuación de las organizaciones.

El marco permite hacer un estudio exploratorio de la influencia que juegan las evaluaciones en la práctica de las organizaciones, y a la vez, contrastar el fenómeno en los distintos tipos de organización de forma que se identifiquen regularidades y diferencias.

2.2.6 Diseño de instrumentos

El marco conceptual guio el diseño de los instrumentos de recolección de información. A partir de las variables se desarrollaron ítems para el guion de la entrevista semi-estructurada, los cuales orientaron a su vez las observaciones y la revisión documental (Anexo 1).

2.2.7 Selección de estudios de caso

En la investigación, la unidad de análisis es la organización, por lo que un caso se refiere a una organización. Para seleccionar los casos, se valoró su relevancia en función del propósito del estudio. Un requisito era que la organización condujera proyectos de desarrollo relacionados con el manejo de recursos naturales dentro de la Reserva. Por otro lado, dado que se buscaba registrar las diferencias entre los casos y cubrir una amplia gama de configuraciones organizacionales, otro requisito era que los casos representaran organizaciones con diferentes tipos de estructura operativa, sistemas de soporte y arreglos institucionales.

Es algo evidente en sí mismo que las investigaciones empíricas en ciencias sociales dependen de la participación de los actores estudiados. Sin su cooperación, no es posible llevar a cabo el estudio, o bien, afecta de forma importante la credibilidad de una investigación. Por ello, el acceso a la información³³ fue otro criterio de selección utilizado.

Sin embargo, dado que la información sobre las evaluaciones es especialmente delicada, no siempre es bienvenida la mirada externa de un investigador en torno del tema. Las evaluaciones generan ansiedad pues se relacionan con la credibilidad de las organizaciones, y el personal de estas podrían temer, no sin razón, que la difusión de información podría amenazarles (Donaldson *et al.*, 2002; McKinnon *et al.*, 2015). Esto se exagera en el campo del desarrollo, ámbito en el que lo sagrado de la misión de reducir la pobreza puede obstaculizar intentos que exploran fallas o incongruencias. Lo anterior, por tanto, puede provocar la resistencia hacia esfuerzos externos de análisis que sean percibidos como un riesgo.

Por ello, para que los integrantes de una organización puedan involucrarse en la reflexión sobre su práctica, es necesario que existan altos niveles de confianza, de forma que sea seguro compartir abiertamente su opinión y experiencia (Serrat, 2010; Kusters *et al.*, 2011). En este contexto, se tomó la decisión de garantizar el anonimato de las organizaciones que participaron en la investigación. Como consecuencia, no se brinda información específica sobre la reserva de la biosfera en la que trabajan.

Lo anterior permite, además, cumplir con estándares éticos de investigación en ciencias sociales. Por un lado, al solicitar el consentimiento informado de los participantes cuando se recaba la información, y por el otro, al respetar sus derechos, intereses y privacidad cuando se divulgan los resultados. Si bien, normalmente se piensa en grupos vulnerables y no en organizaciones cuando se habla de estos estándares, es un tema abierto a discusión en la etnografía institucional la pregunta de si los empleados de organizaciones públicas deben tener las mismas protecciones que se ofrecen a otros sujetos de investigación (Devi, 2005; Dixon y Woods, 2011).

La reflexión ética de las implicaciones de estudiar organizaciones públicas ha generado diversos puntos de vista sobre cómo abordar el dilema del nivel de protección ofrecido. Por un lado, podría argumentarse que, por motivos de transparencia, el personal de organizaciones públicas tiene la obligación de permitir ser estudiados. Así, desde una perspectiva utilitarista, el valor de democratizar el conocimiento justificaría el que los empleados toleren los riesgos e incomodidades asociados al ser estudiados (Dixon y Woods, 2011).

Por otro lado, desde una perspectiva deontológica, asociada a la obligación de no generar daños, es necesario garantizar el derecho a la protección y a la autonomía al personal de las organizaciones participantes. Por ello, se suele usar como norma ética la confidencialidad, incluyendo en ello el no nombrar el sitio de estudio para

³³ El acceso a la información se comprende como la suma de la disponibilidad de información y la voluntad para compartirla en las organizaciones.

prevenir que las personas e organizaciones puedan ser identificadas por otros. En el fondo, son los comportamientos y experiencias los que son de interés científico y no exponer a individuos. Sin embargo, esta decisión acarrea limitaciones en estudios cualitativos. Al no brindarse detalles específicos, los casos quedan con contextos bosquejados que limitan la comprensión del fenómeno. Por ende, se pierde buena parte de la riqueza de los estudios de caso (Walford, 2006).

Lamentablemente, la perspectiva utilitarista también enfrenta problemas epistemológicos en torno de temas delicados como las evaluaciones. Como ya ha sido expuesto, abordajes francos en los que los científicos sociales declaren abiertamente su intención de hacer pública toda la información obtenida podrían enfrentar la resistencia de las organizaciones. El personal estudiado tiene múltiples formas de evadir el escrutinio, negar el acceso, o frustrar intentos de recolección de información. Si, para evitar estos problemas el investigador no es sincero sobre su objetivo de investigación y se acerca a las organizaciones ocultando sus verdaderas intenciones, entonces infringe códigos de conducta profesional por más que se justifique diciendo que lo hace por el bien público. Como se ve por lo expuesto, no hay respuesta sencilla en el establecimiento del balance adecuado entre transparencia y la protección de participantes.

El dilema asociado a la confidencialidad en el manejo de la información tiene un nivel de dificultad adicional cuando el investigador trabaja y prevé trabajar en el largo plazo con las organizaciones estudiadas. En esos casos, la posibilidad de romper la confianza es una implicación a anticipar y evitar, dada la evidencia de que los estudiados podrían experimentar un sentimiento de traición aunque consideren que la exposición de la situación es acertada (Eyben, 2009).

No es un lugar cómodo el usar simultáneamente dos sombreros, el de antropólogo crítico y el de colaborador empático involucrado en la mejora de las organizaciones. Al ocupar esa posición liminal, se corre el riesgo de perder la credibilidad en ambas esferas, tanto ante la academia como con el personal de las organizaciones.

Sin embargo, ubicarse en ese umbral me parece la posición ética más adecuada si se desea jugar un rol crítico constructivo al trabajar con el personal de las organizaciones estudiadas. Por un lado, brinda un espacio de confianza para hablar abiertamente de temas delicados y plantear mecanismos de mejora; por el otro, externa el análisis impersonal de instituciones e incentivos, el cual puede resonar y ser de utilidad para los lectores externos de la investigación.

Se tuvo contacto con tres organizaciones que no fueron incorporadas en el estudio. En dos casos (una dependencia del gobierno federal y una OSC) no fue posible establecer una relación más allá de la entrevista inicial. En la dependencia, no fueron respondidos dos correos de seguimiento para agendar encuentros; en la OSC, una asociación ganadera, se comentó al investigador que no consideraban que su organización aportaría información de interés y declinaron mantener más entrevistas. En el tercer caso, una OSC, si bien existió apertura para participar en el estudio y se realizaron algunas entrevistas, la organización estaba implementando exclusivamente un proyecto multi-anual fuera del área de la reserva.

Aun así, el acercamiento a todas ellas sirvió para probar algunos supuestos en los instrumentos y para refinar la estrategia de acercamiento.

Al final, se seleccionaron cuatro casos, dos centrales, en los que fue posible realizar descripciones y análisis más detallados, y dos secundarios, con la finalidad de establecer comparaciones en torno de variables selectas. El grado de profundidad en el estudio de los casos tuvo relación con la capacidad del investigador para acceder a la información, pero también con la disponibilidad de recursos, pues no era factible lograr el mismo nivel de involucramiento con todas las organizaciones. Las organizaciones estudiadas en los casos centrales fueron una universidad pública y una dependencia federal que tiene una dirección en la reserva: la Comisión Nacional de Áreas Naturales Protegidas (CONANP). Las organizaciones de los casos secundarios fueron una administración municipal y la OSC de mayor tamaño con presencia en la reserva (Cuadro 22).

La decisión de hacer una distinción en el tratamiento de los casos se debió a la relevancia de profundizar en el análisis cuando fuera posible y al interés del investigador por involucrarse en el desarrollo de capacidades en evaluación si se presentaba la oportunidad. Aun así, se encontró una coyuntura para reunir a las cuatro organizaciones (y a otras más) en un foro donde se tematizó la importancia de las evaluaciones en el manejo de la reserva.

Cuadro 22. Relación de organizaciones seleccionadas en comparación con el total de organizaciones similares con presencia en la reserva.

Organización seleccionada	Categoría	Número de organizaciones similares presentes en la reserva
Universidad pública	Caso central	1
Dependencia federal	Caso central	2
Administración municipal	Caso secundario	Más de 3
OSC	Caso secundario	Más de 3

Fuente: Elaboración propia

2.2.8 Preparación del acercamiento a las organizaciones

Como se ha dicho, el acceso a la información era el principal desafío de la investigación. Por ello, se prestó especial atención en la estrategia para construir

una relación empática³⁴ y de confianza con las personas que laboran en las organizaciones estudiadas. Dadas las dificultades que proyectos similares han enfrentado, son pocas las precauciones que se pueden tener para evitar que la investigación sea percibida como una amenaza (Taut, 2007; Eyben *et al.*, 2007; Okail, 2009; Shulha *et al.*, 2016).

Como primer paso, se obtuvo información indirecta sobre organizaciones que implementaban proyectos de manejo de recursos naturales en la Reserva (perfil de la organización, características de los proyectos y del personal). Además, cuando fue posible por las relaciones preexistentes del investigador, se mantuvieron entrevistas con profesionistas que conocían el trabajo de las organizaciones. El objetivo era doble: tener una visión panorámica sobre el uso de evaluaciones que permitiera refinar los instrumentos y aprovechar relaciones que pudieran facilitar el acceso a integrantes de las organizaciones.

2.2.9 Acercamiento a las organizaciones

Se concertó una cita con directivos de cada una de las organizaciones para presentar los objetivos del proyecto y solicitar su consentimiento para interactuar con personal de la organización. En los casos secundarios se solicitó autorización para hacer entrevistas y acceder a documentos de la organización. Mientras tanto, en los casos centrales, se hizo la propuesta adicional de apoyar a la organización en la mejora de su sistema de evaluación (Anexo 2).

El contexto de las organizaciones es tan diverso, que no puede existir una guía sobre cómo llevar a la práctica la PdA propuesta por Eyben. Investigadores vinculados con el desarrollo de capacidades en evaluación proponen comenzar por la recolección de información sobre las organizaciones con las que se pretende colaborar (Cuadro 23). Esto permite conocer la forma en la que hacen y usan evaluaciones, valorar su capacidad para hacerlas, e identificar barreras para el cambio en la escala organizacional y normativa (Preskill y Boyle, 2008; King y Stevahn, 2013).

³⁴ Al hacer evidente el respeto por el trabajo de la organización y la comprensión de los dilemas a los que se enfrenta quién promueve proyectos de desarrollo.

Cuadro 23. Marco para valorar si es posible interactuar con las organizaciones para promover cambios en su sistema de evaluación.

PUNTOS A CONSIDERAR	POSIBLES FALLOS
<ul style="list-style-type: none"> - Conocer la disponibilidad de tiempo y recursos que podrían destinar. - Identificar qué propósito tienen las evaluaciones. - Conocer la resistencia que podrían tener hacia el tema de las evaluaciones o hacia el cambio. - Conocer si habría interés del personal por participar y qué tipo de incentivos podrían utilizarse. 	<ul style="list-style-type: none"> - No hay confianza hacia el proceso. - Los directivos o tomadores de decisión en otro nivel no están interesados. - No pueden cambiar la forma en que trabajan. - No hay tiempo disponible. - No hacen y no les interesan las evaluaciones.

Fuente: Adaptado de King y Stevahn (2013)

Esta información permitiría reconocer el grado de alineación e interés que tienen las organizaciones hacia la propuesta. Mientras que la alineación valora si el punto de vista de la organización coincide con el del investigador, el interés explora si la organización desea que ocurra un cambio, y si dedicaría tiempo y recursos para que suceda (Mendizabal, 2010).

Evidentemente, es necesario reconocer si el personal de las organizaciones puede actuar de una manera diferente a como lo hace. Con esa base, se puede explorar la modificación de algunos aspectos en el corto plazo, los cuales permitan mejorar la práctica evaluativa de las organizaciones.

Dada su complejidad, es incierto si investigaciones realizadas desde una perspectiva de aprendizaje ocasionarán el cambio esperado. Por ello, puede estudiarse también el proceso a través del cual el investigador se involucró con las organizaciones. De esa manera, la reflexividad se vuelca sobre la investigación misma, para analizar los factores que influyeron en el fomento de los cambios deseados.

2.2.10 Estudio piloto

Con la finalidad de probar los instrumentos, se hicieron entrevistas a profesionistas que tienen información sobre la operación de las organizaciones estudiadas, en todos los casos, personas que laboraron previamente en ellas. Además, se aprovechó la interacción con las dos organizaciones que al final no se incluyeron en la investigación.

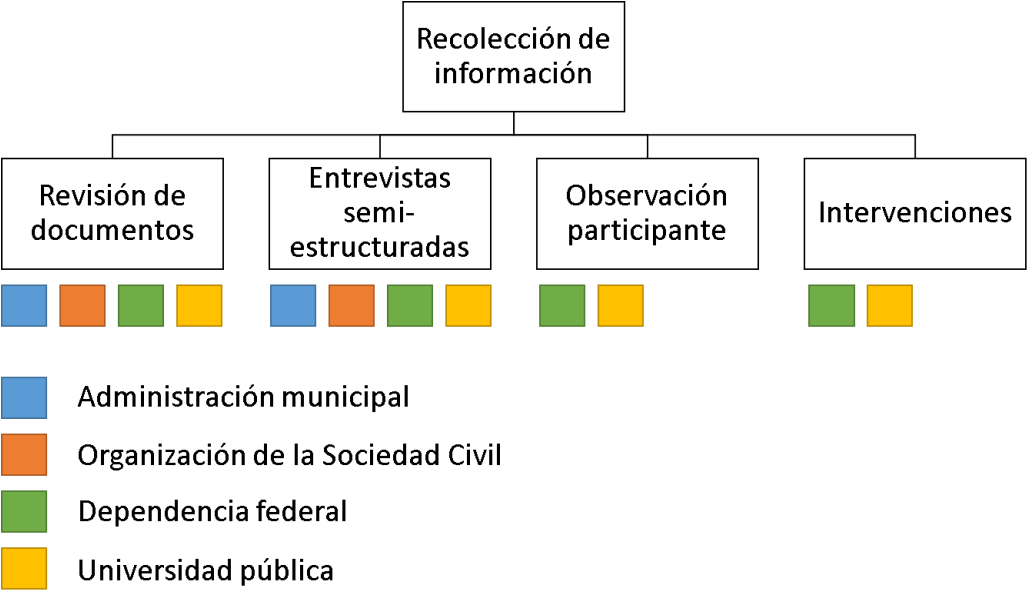
Gracias a lo anterior, se detectaron una serie de ítems cuyo grado de sofisticación no correspondía con la práctica de las organizaciones estudiadas. Por ello, la respuesta obtenida a preguntas básicas develó que algunos ítems del instrumento asumían erróneamente la presencia de ciertos elementos en las organizaciones. Esto permitió simplificar el instrumento, reduciendo el número de variables consideradas del marco conceptual.

A su vez, se identificaron las limitaciones de la entrevista semi-estructurada para medir algunos ítems relacionados con la cultura organizacional. Por esos motivos, se tomó la decisión de ceñir el instrumento a la medición de manifestaciones visibles del comportamiento organizacional.

2.2.11 Recolección de la información

El estudio recurrió al análisis de documentos, entrevistas semiestructuradas, observación participante e intervenciones para reunir la información necesaria y responder así a las preguntas de investigación (Figura 9).

Figura 9. Técnicas de recolección de la información.



Fuente: Elaboración propia

Revisión de documentos

En cada caso, la recopilación de documentos relacionados con las evaluaciones se llevó a cabo a través de una exploración de sus páginas web institucionales, y a

través de solicitudes directas al personal directivo de la organización. De esa manera se tuvo acceso a evaluaciones externas, informes, y documentos de planeación estratégica. También se recopilaron datos secundarios a través de una búsqueda de artículos y tesis que han abordado el tema de las evaluaciones en la Reserva.

Entrevistas semi-estructuradas

Las entrevistas permitieron recolectar información sobre la práctica de las organizaciones, entremezclada con la opinión y reflexiones de los entrevistados en torno del tema. La selección de las entrevistas dentro de las organizaciones se hizo a personal directivo y a personal operativo, con base en el juicio personal del investigador. El principal criterio de selección fue la relación con procedimientos relacionados con evaluaciones. Sin embargo, la selección tomó en cuenta la inclusión de actores que se encuentran en diferentes posiciones, dada la posible existencia de diferencias intra-organizacionales.

En total, se llevaron a cabo 29 entrevistas semi-estructuradas, once a personal de la dependencia federal, diez a personal de la Universidad Pública, cuatro a personal de la OSC y cuatro a personal de la administración municipal (Cuadro 24).

Cuadro 24. Número de entrevistas y de documentos revisados.

Organización	Número de entrevistas	Número de documentos revisados
Dependencia federal	11	35
Universidad pública	10	22
Administración municipal	4	17
OSC	4	24

Fuente: Elaboración propia

Observación participante

A través de la observación participante, fue posible explorar la cotidianidad de las organizaciones al interactuar con su personal. Esto permitió hacer descripciones detalladas de diferentes aspectos relacionados con su cultura organizacional. La técnica sólo se pudo emplear con dos organizaciones, la dependencia federal y la universidad pública.

De manera específica, se participó en tres salidas de campo con personal de la dependencia, dos de trabajo con comunidades y una de monitoreo biológico.

Además, se asistió a varias reuniones del consejo asesor de la reserva, órgano de participación promovido por la CONANP. Por su parte, se asistió a múltiples reuniones trabajo de la universidad pública, algunas de ellas facilitadas por el investigador.

Intervenciones

Debido a que se adoptó la PdA, una de las intenciones de la investigación fue buscar abiertamente coyunturas en las que fuera posible realizar observaciones por medio de una intervención. El propósito central de las intervenciones, en ese sentido, sería facilitar un espacio de diálogo que permitiera generar conocimientos y promover un proceso de cambio en las organizaciones participantes.

La existencia de esas oportunidades no podía preverse de antemano, y podía ser el caso que el investigador no tuviera la capacidad y/o la fortuna de impulsar el proceso en las organizaciones. Para que la vinculación pudiera concretarse, no solo tenía que evitarse que la investigación fuera percibida como una amenaza, sino que tendría que considerarse como algo útil a los intereses de las organizaciones.

Fue posible realizar tres intervenciones, una conectada con intereses de la CONANP, pero que involucró al resto de las organizaciones, y las otras dos con personal de la universidad pública. Se describen a continuación.

1. Foro “Evaluación de la efectividad en el manejo de la Reserva”

Durante la investigación, el equipo de guardaparques se encontraba en el proceso de actualización del programa de manejo de la reserva, documento de planeación de gran importancia tanto para ellos como para otras organizaciones en la región por sus implicaciones legales.

En ese contexto, el investigador organizó y facilitó, con el respaldo de los guardaparques, un foro para promover conversaciones en torno de la importancia de las evaluaciones para el manejo efectivo de la reserva. En él participaron 34 personas provenientes de dos administraciones municipales, tres dependencias de gobierno, tres OSC y dos universidades públicas.

A través de técnicas de liderazgo participativo como el café mundial (Brown e Isaacs, 2008), el espacio abierto (Owen, 2008) y la indagación apreciativa (Catsambas, 2016), el foro promovió la discusión de ventajas, retos y posibles fallas de alianzas para evaluar la efectividad en el manejo de la reserva.

Algunas de las preguntas convocantes fueron:

- ¿Qué indicadores biofísicos, sociales e institucionales podemos incorporar para tomar decisiones de manejo?
- ¿Dónde está la intersección entre los incentivos y los procedimientos de nuestras organizaciones para desarrollar una alianza exitosa?

- ¿Cómo institucionalizar el aprendizaje conjunto para ser cada vez más efectivos?

La sistematización de los resultados del taller y las estrategias de seguimiento se presentan en el anexo 3.

2. Ruta de generación de impacto de un proyecto de conservación *ex situ*

Se aprovechó la invitación para colaborar en un proyecto del centro de investigación agrícola de la universidad pública, a punto de iniciar y ya financiado, para ofertar una evaluación de diseño que mejorara la descripción de sus impactos. El resultado se presenta en el capítulo 5.

3. Propuesta de reforma al sistema de evaluación científica de la Universidad Pública

Al mantener contacto con autoridades de la universidad pública, el investigador supo que la administración central estaba buscando nuevas formas de evaluar el desempeño global de las facultades. El propósito del proyecto es incentivar a su personal a través de un esquema similar al Presupuesto basado en Resultados.

En esa coyuntura, se ofertó a la Dirección de Investigación el organizar y facilitar una serie de mesas redondas para discutir desafíos y áreas de oportunidad en el sistema de evaluación de las investigaciones aplicadas empleado por la universidad pública. En él participarían investigadores y jefes de investigación de cada facultad. El evento ya había sido aceptado para realizarse en el marco de la Semana de la Evaluación 2020³⁵, pero tuvo que cancelarse debido a la contingencia sanitaria por COVID-19. Ante esta situación, y aprovechando la reelección de las autoridades universitarias para el periodo 2021-2024, se integró una propuesta de continuidad consistente en la reforma del sistema de evaluación científica de la universidad pública. La información se presenta en el anexo 4.

2.2.12 Limitaciones de la metodología utilizada

Las limitaciones del abordaje metodológico son las propias de cualquier investigación exploratoria de un problema complejo. No se puede esperar como resultado la identificación de relaciones causales entre las variables, pero sí la detección de patrones que permitan cuestionar supuestos existentes y generar nuevas hipótesis.

Por la magnitud del estudio, fue posible obtener evidencia empírica para:

- caracterizar y comparar la forma en la que las organizaciones hacen y usan evaluaciones,

³⁵ <https://clear-lac.org/semana-de-la-evaluacion/>

- identificar las fortalezas, retos y limitaciones de sus sistemas de evaluación.

Sin embargo, en un sólo estudio de caso hay mucho por cubrir, por lo que se necesitaría un análisis de mayor profundidad para cubrir el espectro completo del fenómeno a estudiar. En ese sentido, al abarcarse tantos casos, algunas descripciones y análisis quedaron a un nivel superficial, especialmente en los casos secundarios en torno de variables difíciles de medir como el liderazgo, la cultura organizacional y los efectos de las evaluaciones sobre el comportamiento del personal de la organización. Esto es patente en el caso de la OSC, en donde a pesar de ser una de las organizaciones con mayor cantidad de evaluaciones, quedó débilmente descrita la influencia de estas sobre su toma de decisiones.

En un sentido similar, la carga de trabajo del investigador le impidió dar el seguimiento que hubiera querido a la intervención realizada con la CONANP. Si bien se diseñó una página web para compartir los resultados del foro, y además se construyó un grupo de google para facilitar la colaboración de aquellos participantes interesados en el nuevo programa de manejo de la reserva, la propuesta de organización no tuvo la efectividad esperada.

Capítulo tres

Contexto del estudio de caso

Como consecuencia de la decisión de mantener el anonimato tanto de las organizaciones como de la reserva en donde laboran, se ubica el área de estudio en un marco de referencia más amplio: la región administrativa Centro y Eje Neovolcánico establecida por la CONANP. Después, se presentan rasgos de la reserva que permiten comprender su dinámica ambiental; sin embargo, no se citan referencias que permitan la identificación de la reserva. Finalmente, se describen características de cada una de las organizaciones estudiadas y se perfilan elementos para comprender sus sistemas de evaluación, de nuevo, sin referencias específicas.

3.1 REGIÓN CENTRO Y EJE NEOVOLCÁNICO

México es uno de los países más biodiversos a nivel mundial, contribuyendo con el 10% de la riqueza biológica global de cada taxón de organismos (Mittermeier y Mittermeier, 1997; Llorente y Ocegueda, 2008). Históricamente, el establecimiento de ANP ha sido uno de los instrumentos de política pública empleados para conservar ese patrimonio. La Comisión Nacional de Áreas Naturales Protegidas (CONANP) es, desde el año 2000, la autoridad federal encargada de administrarlas (CONABIO, 2017).

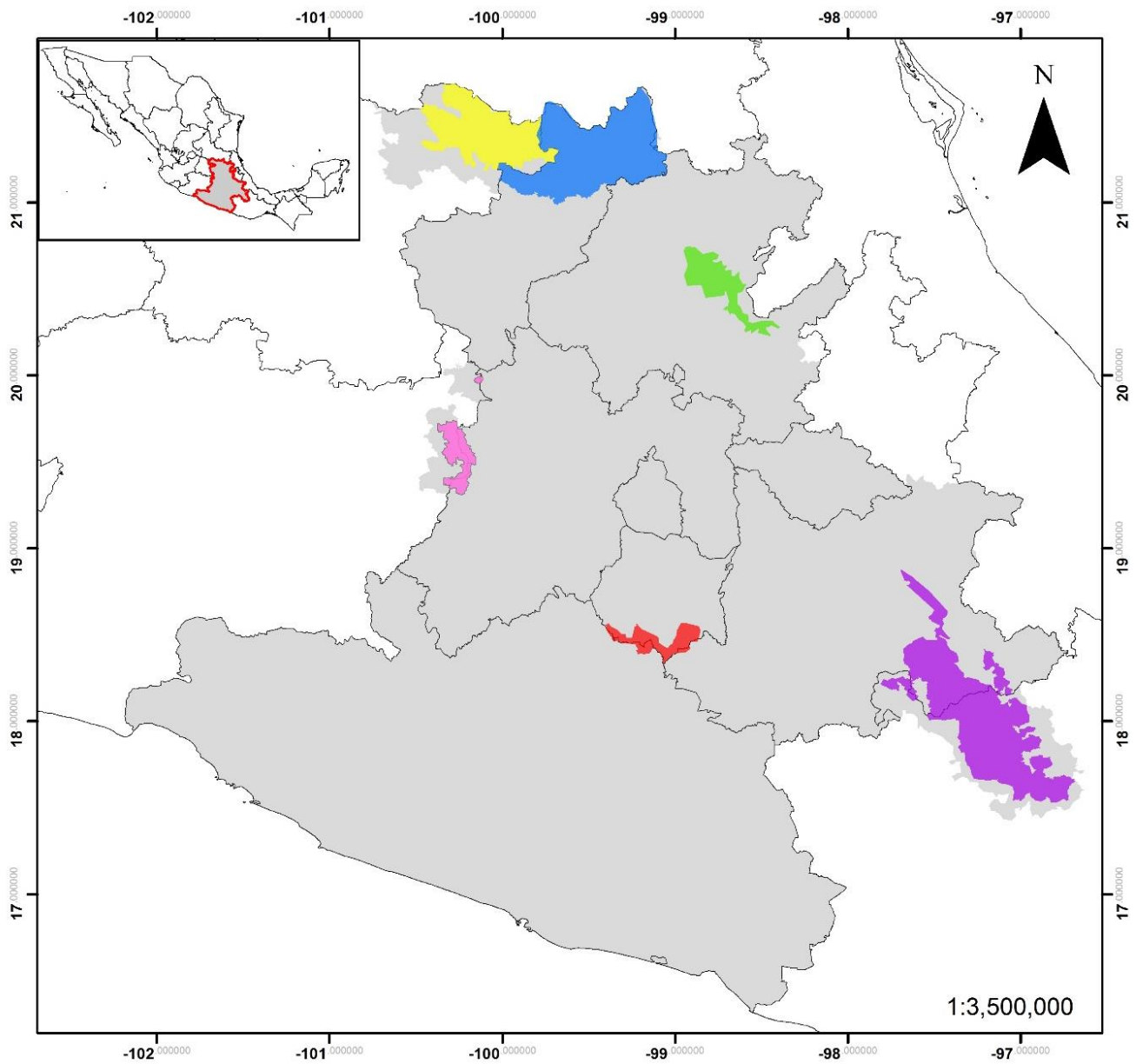
Para manejar las ANP a nivel federal, la CONANP dividió el país en nueve Direcciones Regionales. El área de estudio de la investigación es una de las seis reservas de biosfera ubicadas en la región Centro y Eje Neovolcánico (Mapa 1), la cual abarca una superficie cercana a 16,500,000 hectáreas en once estados de la república³⁶, y forma parte de siete cuencas hidrográficas: Balsas, Costa Chica de Guerrero, Costa Grande de Guerrero, Lerma-Santiago, Norte de Veracruz, Pánuco y Papaloapan (CONANP-GIZ, 2018a).

El área es la más densamente poblada de México. Por ello, dada la cercanía de las ANP de la región a muchos centros urbanos, existe dependencia hacia los servicios ecosistémicos que éstas proveen, principalmente agua para consumo humano (CONANP-GIZ, 2018b).

Debido a la sobreposición de las ecozonas Neártica y Neotropical, es una de las regiones con mayor riqueza de especies y endemismos (CONANP-GIZ, 2018b). En ella coinciden, además, una gran diversidad de ecosistemas: zacatonal de altura, bosque de coníferas, bosque de encinos, bosque mesófilo de montaña, selva seca, selva tropical, matorrales y humedales (Figura 10).

³⁶ Ciudad de México, Estado de México, Guanajuato, Guerrero, Hidalgo, Michoacán de Ocampo, Morelos, Oaxaca, Puebla, Querétaro y Tlaxcala.

Mapa 1. Reservas de la Biosfera ubicadas dentro de la Dirección Regional Centro y Eje Neovolcánico











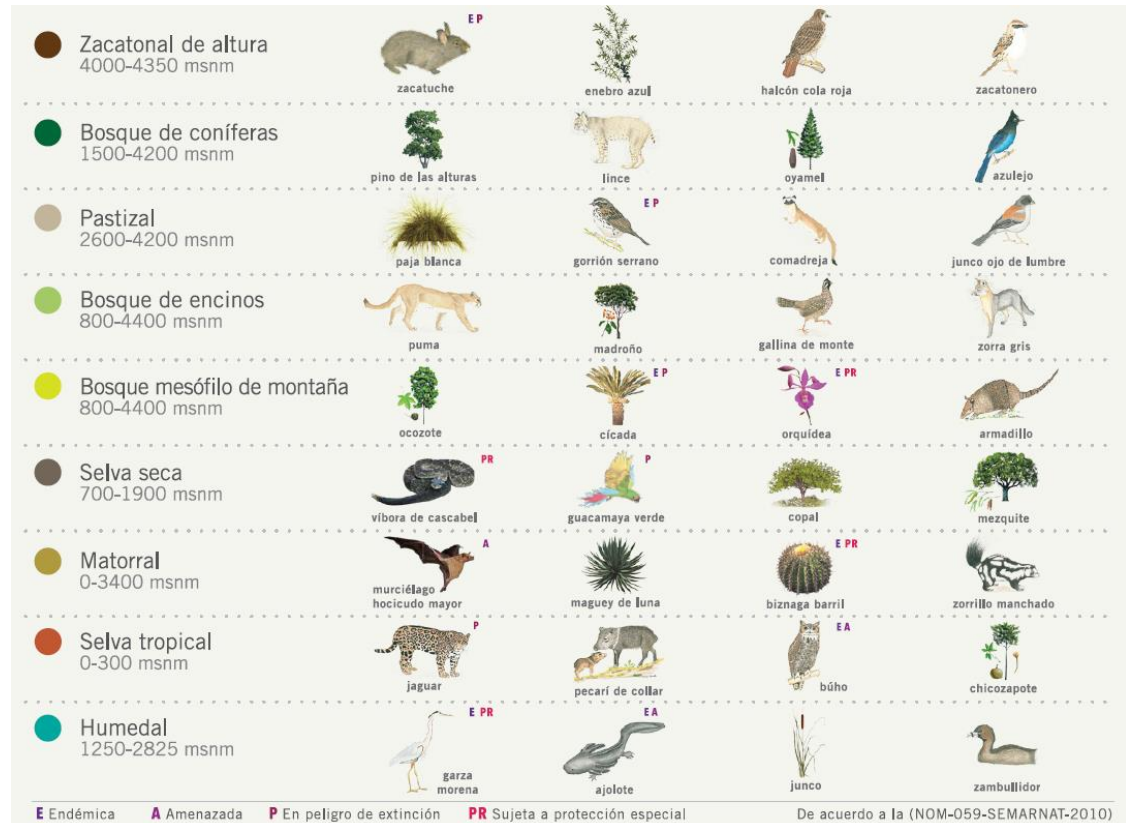
 Dirección Regional Centro y Eje Neovolcánico	
RESERVAS DE LA BIOSFERA	Fuentes cartográficas
 Sierra Gorda de Guanajuato	- CONANP, 2017. Regiones CONANP.
 Sierra Gorda de Querétaro	- CONANP, 2019. Áreas Naturales Protegidas
 Barranca de Metztitlán	Especificaciones del mapa
 Mariposa Monarca	Proyección..... WGS84
 Sierra de Huautla	Datum..... World Geodetic System 1984
 Tehuacán-Cuicatlán	Fecha..... Noviembre, 2020

Figura 10. Especies representativas y ecosistemas de la región Centro y Eje Neovolcánico



Fuente: CONANP-GIZ, 2016



En la región 1,321,971 de las 1,729,164 hectáreas bajo régimen de protección han sido declaradas como Reservas de la Biosfera (76%), categoría que implica la búsqueda simultánea de la conservación y el desarrollo sostenible de la sociedad rural que habita esos espacios. Esto es congruente, dado que en sus seis reservas de la biosfera se asientan cerca de 1,500 localidades, varias de ellas indígenas³⁷, con una población total que supera los 200,000 habitantes (CONANP-GIZ, 2018a).

3.2 CARACTERÍSTICAS GENERALES DE LA RESERVA DE LA BIOSFERA

La zona fue declarada área protegida debido a la gran diversidad de especies y ecosistemas que aloja. Además, se ha resaltado su importancia hidrológica, pues sus bosques juegan un rol fundamental en la provisión de agua al facilitar la infiltración y la recarga de acuíferos. Sin embargo, a pesar de su carácter estratégico para la seguridad hídrica y la conservación de la biodiversidad, se han documentado diversas problemáticas ambientales como la ganadería extensiva que afecta la regeneración natural de los bosques, plagas forestales y la propagación de especies acuáticas invasoras.

En torno del cambio en el uso del suelo, uno de los principales indicadores de degradación/restauración ecológica utilizados, existen datos contrastantes. Un estudio determinó, a través de la interpretación de imágenes satelitales, una recuperación de áreas con vegetación en el periodo 1973-2004. Los autores concluyeron que, si bien podían existir errores metodológicos, una hipótesis plausible de la recuperación sería un proceso de abandono de áreas agrícolas ocasionado por la migración. Por otro lado, entre 2000 y 2010, otros estudios valoraron la efectividad de la reserva para contener procesos de cambio en el uso del suelo al hacer una comparación de la tasa de cambio en la reserva y su contexto inmediato. La reserva se consideró no efectiva, estimándose una pérdida de 2,775 hectáreas de suelo forestal.

En la reserva hay grandes contrastes sociales, pues las zonas forestales donde habita la población indígena y campesina tienen muy bajos niveles de desarrollo social y económico. Por ello, la migración regional e internacional ha sido una estrategia histórica para conseguir la manutención familiar. Este proceso es evidente en el limitado crecimiento poblacional de las últimas décadas, el índice de intensidad migratoria de los municipios y el porcentaje de viviendas que reciben remesas.

Además, los efectos de la alta migración masculina son perceptibles en dos esferas. Por un lado, en la estructura familiar y comunitaria, ha incrementado la participación de las mujeres en la vida pública de las comunidades. Por el otro, se ha fortalecido

³⁷ En las reservas habitan comunidades de los pueblos indígenas Pame, Otomí, Mazahua, Náhuatl, Teenek, Chinanteco, Chocholteco, Ixcateco, Mazateco, Mixteco y Popoloca.

un proceso de desagrarización de la economía local, reduciéndose las expectativas de vivir de actividades agropecuarias o forestales en el mediano plazo.

Aun así, en la región se sigue practicando la agricultura de temporal, dedicada principalmente a la producción de maíz y frijol para el autoconsumo. La actividad se desarrolla principalmente en minifundios sobre pendientes medias y altas (mayores al 20%), por lo que tienen una baja aptitud agrícola y una alta susceptibilidad a la erosión. Por su parte, la ganadería se orienta a la producción de pies de cría de bovinos bajo un sistema extensivo de libre pastoreo, con el objetivo de abastecer el mercado local o de servir para el autoconsumo. La escasa disponibilidad de forraje, principal problemática señalada por los ganaderos en la región, impulsa el establecimiento de pastizales inducidos a costa de áreas forestales, siendo esta la principal causa del cambio de uso de suelo.

En la región, poca superficie forestal es apta para el aprovechamiento maderable. Si bien existen algunos aprovechamientos autorizados, la finalidad principal de la actividad es abastecer de leña y materiales de construcción a las comunidades en el mercado local, en la mayoría de los casos de forma clandestina. En ciertas áreas, la extracción ha sido superior a la capacidad de regeneración natural, generando un proceso de fragmentación de ecosistemas forestales.

En los últimos años se ha incrementado de forma importante el flujo de turistas a la reserva, lo cual ha permitido el desarrollo de negocios ligados a la oferta de servicios. Hacen falta estudios sobre su impacto global, sin embargo, se han advertido riesgos vinculados con una baja regulación y vigilancia, principalmente en parajes naturales que podrían ver rebasada su capacidad de carga.

Actualmente, distintas dependencias de gobierno, instituciones académicas y organizaciones de la sociedad civil llevan a cabo proyectos que buscan cumplir de forma simultánea con objetivos sociales, económicos y ambientales. Sin embargo, los proyectos de desarrollo han tenido baja efectividad a escala regional, pues sus efectos acumulativos no han revertido las problemáticas mencionadas.

En el siguiente apartado, se presentan las características de las organizaciones estudiadas, se especifica el propósito de sus proyectos ligados al manejo de recursos naturales y se desarrollan elementos básicos de su sistema de evaluación.

3.3 ORGANIZACIONES ESTUDIADAS

3.3.1 Dependencia federal

La dependencia federal estudiada fue la Comisión Nacional de Áreas Naturales Protegidas (CONANP), órgano descentralizado de la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). Esta entidad es responsable de conservar la biodiversidad en los ecosistemas más representativos de México.

A escala nacional, la CONANP se apoya en nueve direcciones regionales para administrar 182 ANP que abarcan más de 90 millones de hectáreas (CONANP-GIZ, 2018a). Por su parte, a escala local, direcciones de ANP en las que laboran equipos de guardaparques son las directamente encargadas de manejar las áreas decretadas. Para ello, implementan líneas de acción que parten tanto de directrices institucionales (Cuadro 25) como de un programa de manejo diseñado para enfrentar las problemáticas del ANP (CONANP, 2014c).

Cuadro 25. Ejes sustantivos de la estrategia 2040 de la CONANP.

Eje	Objetivo
Manejo integrado del Paisaje	Lograr un manejo integrado del paisaje de las ANP y sus zonas de influencia para asegurar la conservación de la biodiversidad y de los servicios ambientales generados.
Conservación y manejo de la biodiversidad	Mantener la representatividad de la biodiversidad, la conectividad y función de los ecosistemas y la integralidad de los procesos ecológicos.
Participación social y cultura para la conservación	Lograr que la sociedad valore, sea corresponsable y participe activamente en el manejo sustentable de la biodiversidad y servicios ecosistémicos.
Economía de la conservación	Contribuir al fortalecimiento de la economía desde un enfoque regional, promoviendo la valorización de los bienes y servicios ecosistémicos para impulsar actividades productivas y negocios sustentables.
Cambio climático	Contribuir a incrementar la resiliencia de las comunidades humanas, los ecosistemas y sus servicios ambientales, así como a la mitigación del cambio climático en las ANP.

Fuente: CONANP, 2014c

De esta forma, el objetivo de las Direcciones de Área Natural Protegida es mantener la continuidad de los procesos ecológicos y, a la vez, lograr un desarrollo económico y social para los habitantes del polígono señalado por la declaratoria de ANP (CONANP, 2014a). Algunas de sus funciones son:

- Administrar, manejar y ejecutar las acciones para conservar los ecosistemas y su biodiversidad.
- Formular y ejecutar el programa de manejo correspondiente, así como proveer los elementos necesarios para su evaluación y modificación.
- Conformar el sistema de información con los datos biológicos, sociales, económicos y cartográficos del ANP.

- Fomentar la celebración de instrumentos jurídicos con los sectores público, social y privado para ejercer acciones de administración, conservación, mejoramiento e investigación de los ecosistemas.

En la organización, el déficit financiero para manejar de forma efectiva las ANP es uno de los asuntos más acuciantes. Aunque en los últimos cinco años el número de hectáreas salvaguardadas se ha incrementado 2.5 veces, el presupuesto de la dependencia se ha reducido en 47.3% (Plaza, 2020). Esta brecha presupuestal limita, en definitiva, la operación de los programas de trabajo de los guardaparques.

En la reserva, la dirección de ANP está conformada por diez guardaparques, siendo liderados por un director y sub-director de área. En el equipo prevalecen perfiles profesionales asociados con el manejo de recursos naturales (biología, ingeniería forestal, veterinaria), pero también hay profesiones ligadas a áreas administrativas (administración, turismo). Las principales actividades de los guardaparques se relacionan con el monitoreo biológico, la educación ambiental, el control de incendios y la coordinación de los programas de subsidio de la dependencia.

A escala nacional, el desempeño de la dependencia es evaluado tanto a través de la verificación del cumplimiento de los objetivos establecidos en el Programa Nacional de Áreas Naturales Protegidas (Cuadro 26) como por medio de evaluaciones externas de sus programas de subsidio, las cuales buscan apoyar la toma de decisiones presupuestales y emitir recomendaciones de mejora. Las instancias responsables de coordinarlas son el Consejo Nacional de Evaluación de la Política de Desarrollo Social³⁸ (CONEVAL) y la Auditoría Superior de la Federación³⁹ (ASF). El programa más evaluado de la dependencia ha sido el Programa de Conservación para el Desarrollo Sostenible (PROCOCODES). La ASF le ha hecho dos auditorías de desempeño (2012 y 2016) y la CONEVAL ha efectuado diversos tipos de evaluación a lo largo de nueve años⁴⁰.

Como se ha dicho, sin embargo, la atención de la investigación se coloca en las evaluaciones elaboradas por las propias organizaciones, específicamente en aquellas relacionadas con las decisiones que se toman en la reserva estudiada. Por lo dicho en diferentes documentos institucionales, es evidente que la evaluación es un procedimiento valorado para cumplir metas en la dependencia (CONANP, 2010a, 2010b, 2014a, 2016).

³⁸ La Ley General de Desarrollo Social (2004) establece que el CONEVAL tiene la atribución de evaluar los programas sociales. En 2009 se expidieron los lineamientos generales para la evaluación de los programas federales, los cuales describen los propósitos y características de diferentes tipos de evaluación que se pueden hacer a los programas federales.

³⁹ La Ley de Fiscalización y Rendición de Cuentas de la Federación (2009), establece que la ASF evaluará el desempeño de los programas federales.

⁴⁰ Cinco evaluaciones específicas de desempeño (2008, 2009-2010, 2010-2011; 2011-2012, 2012-2013 y 2014-2015), dos evaluaciones de consistencia y resultados (2007 y 2011), una evaluación de la estrategia de cobertura y focalización (2007-2008), y una ficha de monitoreo (2016-2017).

Cuadro 26. Algunos indicadores de objetivo establecidos en el Programa Nacional de Áreas Naturales Protegidas.

Área	Indicadores	Unidad de medida	Línea base 2013	Meta 2018
Manejo integrado del paisaje	Superficie del territorio nacional dedicada a la conservación a través de ANP.	Hectáreas	25,394,779	54,488,575
Cambio Climático	ANP que incluyen el componente de cambio climático en sus programa de manejo o realizan acciones para el incremento en la captura o almacenaje de carbono.	Número de ANP	23	58
Fortalecimiento institucional	ANP incorporadas a un proceso de evaluación de la efectividad en su manejo.	Número de ANP	0	8
	ANP en que se evalúa la tasa de cambio de las cubiertas del suelo.	Número de ANP	36	62
Economía de la conservación	Empleos generados a través del fomento de actividades productivas sustentables.	Empleos	11,190	58,00
	ANP en que se realiza el cálculo del valor de las actividades económicas.	Número de ANP	0	155
	ANP en que se instrumentan métodos de valoración de servicios ecosistémicos.	Número de ANP	0	10

Fuente: CONANP, 2014

En su estructura operacional, es atribución de la Dirección de Evaluación y Seguimiento proponer las metodologías que se utilizarán a nivel nacional, regional y local. Algunas responden a las necesidades de evaluación antes mencionadas, otras a las necesidades de información de la propia organización y otras a procesos internacionales de rendición de cuentas, como la medición de avances en la meta 11 de Aichi, compromiso adquirido por México ante el Convenio de Diversidad Biológica (CONANP, 2016). Por su parte, es responsabilidad de la Dirección de ANP, no solo llevar a cabo diferentes evaluaciones solicitadas por la Dirección de Evaluación y Seguimiento, sino evaluar el propio programa de manejo del ANP.

3.3.2 Universidad Pública

En México, las Universidades Públicas Estatales (UPE) son instituciones de educación superior que operan bajo la figura jurídica de organismos públicos descentralizados. Para realizar sus funciones básicas (docencia, generación y aplicación del conocimiento), las UPE reciben recursos públicos provenientes de la federación y el estado donde se asientan. Una de sus particularidades es que poseen autonomía institucional, es decir, tienen la facultad y responsabilidad de gobernarse a sí mismas.

La universidad estudiada es la institución educativa con mayor oferta de carreras universitarias en el estado donde se asienta. Además, diferentes rankings que valoran la calidad en la docencia, la producción y transferencia de conocimientos y el reconocimiento internacional la catalogan como una de las mejores del país.

Uno de los objetivos centrales de la organización es realizar investigaciones que tengan proyección social, es decir, que vinculen los conocimientos para resolver problemas y promover el desarrollo del país. Este compromiso social se refleja en varios ámbitos. Por un lado, en su modelo educativo, donde establece como un pilar la responsabilidad de la universidad con las necesidades de transformación de su entorno. Por el otro, en su modelo de gestión organizacional, la Responsabilidad Social Universitaria⁴¹, que promueve el manejo de los impactos generados por la universidad.

A nivel de procedimientos, lo anterior se manifiesta en el reconocimiento de la vinculación como función sustantiva de la universidad. Por ello, la organización cuenta con mecanismos que financian y recompensan el involucramiento de su personal en proyectos que busquen solucionar problemas a través de la vinculación social, tecnológica y artística.

Otra política que expresa el compromiso social de la universidad ha sido el establecimiento de campus regionales en distintos municipios del estado. Si bien la estructura de la organización está basada en facultades, la universidad ha ensayado modelos híbridos basados en la multidisciplinaria para promover el desarrollo local. Ejemplo de ello fue el establecimiento de un centro público de investigación agrícola en la reserva de la biosfera estudiada.

El propósito del centro es generar conocimientos que mejoren los procesos productivos en la reserva a través de la transferencia de tecnología. Además, en él se forman recursos humanos especializados que responden a las necesidades locales. Partiendo de su enfoque pluridisciplinar, en él colaboran diez profesores-investigadores con perfiles provenientes de las ciencias biológicas, agropecuarias y sociales.

⁴¹ El modelo de la Responsabilidad Social Universitaria distingue cuatro tipos de impacto que la universidad puede generar y manejar: organizacionales, educativos, cognitivos y sociales (Vallaeys, 2014).

Los principales proyectos del centro se relacionan con el manejo pecuario (desarrollo de sistemas silvopastoriles, alternativas de nutrición con especies nativas), agrícola (áreas demostrativas y experimentales de producción de hortalizas, plantaciones bioenergéticas), acuícola (sistemas productivos integrando especies nativas), forestal (rehabilitación forestal con especies nativas, manejo integrado de plagas) y financiero (banca comunitaria). En el periodo 2016-2017, para ejecutar varios de estos proyectos, el centro recibió financiamiento proveniente de fondos de la universidad por un monto total de \$600,000.

Aunque la universidad es autónoma, su desempeño es evaluado cuando obtiene fondos extraordinarios concursables provenientes de la federación. Esta gestión es importante, dado el contexto de carestía presupuestal por el que atraviesa la educación superior en México (Mendoza, 2011). Dos fondos importantes durante la última década fueron el Programa Integral de Fortalecimiento Institucional (PIFI)⁴² y el Programa Fortalecimiento de la Calidad Educativa (PFCE)⁴³. Ambos programas partían de indicadores para medir el progreso institucional en torno de la “capacidad y competitividad académica”. En específico, la capacidad académica se vinculaba con el perfil y desempeño de sus investigadores (contar con posgrado, formar parte del padrón de programas de pago por mérito, e integrar cuerpos académicos⁴⁴) (Vidaurri y Luna, 2017).

Independientemente de la evaluación global de la universidad, este tipo de organización interactúa con evaluaciones externas cuando recibe financiamiento para la implementación de proyectos de investigación y vinculación, los cuales pueden ser financiados por entidades públicas o privadas.

Además, es importante mencionar los procedimientos evaluativos que se enfocan directamente en los académicos. Desde los noventa, una serie de programas federales han realizado transferencias monetarias condicionadas para incentivarlos a cumplir con estándares de desempeño, teniendo un peso importante en la estructura de su salario.

Por un lado, el Sistema Nacional de Investigadores (SNI) valora tener el grado de doctor, formar recursos humanos de alto nivel, y tener una alta productividad académica (Gil, 2014). Esta última, medida a través de publicaciones en revistas reconocidas, capítulos de libros, y libros editados o coordinados, es especialmente importante para la permanencia y promoción del académico en las distintas categorías que maneja el SNI (Galaz *et al.*, 2008).

Por el otro, el Programa para el Mejoramiento del Profesorado (PROMEP) y el Programa para el Desarrollo Profesional Docente (PRODEP), buscan profesionalizar a los académicos, de manera que alcancen un “perfil deseable”. Para ello, promueve que formen equipos de trabajo a través de cuerpos académicos.

⁴² Vigente del 2001 al 2013.

⁴³ Vigente desde el 2015.

⁴⁴ Los Cuerpos Académicos son grupos de profesores de tiempo completo que comparten una o varias líneas de generación y aplicación del conocimiento (SEP, 2018).

Estos, a su vez, reciben apoyos financieros para implementar proyectos con el potencial de “transformar su entorno” (SEP, 2018, pág. 16).

En la estructura organizacional, conectando la evaluación externa con la interna se encuentra la Dirección de Planeación. Esta es responsable de recopilar la información de los indicadores con los que se mide el avance en los fondos federales, pero también de coordinar la elaboración de planes institucionales de desarrollo con diferentes horizontes de implementación.

Jerárquicamente hablando, la universidad cuenta con un plan estratégico a 15 años, el cual contiene un conjunto de políticas universitarias con una visión de largo plazo. Después, están los planes institucionales de desarrollo, elaborados cada tres años por la administración central. Estos planes reúnen acciones con metas e indicadores, cuyo cumplimiento es evaluado periódicamente para facilitar la toma de decisiones estratégicas, replantear objetivos y retroalimentar el rumbo de la universidad “en un contexto de manejo adaptativo” (Cuadro 27).

Cuadro 27 Algunos indicadores del Plan Institucional de Desarrollo 2015-2018.

Estrategia	Actividades	Indicador/Meta
Orientación de la investigación a resolver problemas sociales que favorezcan el desarrollo de la región	Establecer convenios de colaboración con los diferentes sectores para el desarrollo de proyectos	Convenios/ Incrementar en 15% el número de convenios
Reconocimiento de la función de vinculación como actividad sustantiva en el quehacer universitario	Incluir en los programas de estímulos y promoción docente actividades de vinculación	Número de profesores reconocidos que participan en proyectos de vinculación / Incremento anual del 5%
	Reconocer a través de distinciones la labor en vinculación	Otorgar un premio universitario de vinculación
Consolidación del Sistema Universitario de Vinculación	Institucionalizar los programas y mecanismos de operación y evaluación de la función de vinculación	Frecuencia de reuniones del consejo de vinculación / recursos destinados al fondo de fortalecimiento a la vinculación
	Establecer procedimientos y lineamientos para el registro y operación de proyectos de vinculación	Lineamientos de operación de proyectos, número de proyectos registrados

Estrategia	Actividades	Indicador/Meta
Evaluación de los planes de desarrollo de los Campi	Desarrollar un instrumento de evaluación de los planes de desarrollo de los Campi	Número de instrumentos elaborados

Finalmente, están los planes de desarrollo de cada una de las facultades y campus que componen la universidad, los cuales precisan elaborarse siguiendo las directrices marcadas en el plan estratégico e institucional. A través de la dirección de planeación, la universidad promovió que el grupo de investigadores que forman parte del centro de investigación agrícola generara el plan de desarrollo del campus donde se encuentra inserto. En él incorporaron acciones específicas para atender problemáticas regionales a través de la investigación.

Así como en la escala organizacional, la Dirección de Planeación es la instancia responsable de establecer indicadores para la evaluación y seguimiento del desarrollo institucional de la universidad, existen otras instancias responsables de procedimientos para evaluar el desempeño de proyectos y del personal.

El diseño de procedimientos para la evaluación *ex ante* y *ex post* de proyectos de investigación es responsabilidad de una Dirección de investigación. En el caso de proyectos de vinculación, esta misma función es cubierta por la Secretaría de Extensión. La dictaminación de los proyectos, tanto para permitir su registro como su finiquito, se basa en la evaluación de pares que verifican el cumplimiento de estándares de calidad y ética. Con respecto de la evaluación del personal, es responsabilidad de la Secretaría Académica coordinar la implementación de evaluaciones del desempeño que permiten tomar decisiones de promoción laboral y contratación.

3.3.3 Administración Municipal

En México, los municipios son la base de la organización política estatal. Por lo previsto en el artículo 115 de la constitución mexicana, son autónomos, es decir, está facultados para definir por sí mismos su estructura organizacional y su normatividad. Los ayuntamientos, encabezados por presidentes municipales, son electos por periodos trianuales, con opción a ser reelectos (CONEVAL, 2016).

El municipio cuya administración se estudió, electo para el periodo 2015-2018, está inserto en su totalidad dentro de la reserva de la biosfera. En su plan de desarrollo municipal, instrumento de planeación que reúne líneas de acción estratégicas, pueden identificarse elementos ligados al manejo de recursos naturales en tres ejes rectores: Apoyo al campo, Desarrollo económico y Servicios Públicos (Cuadro 28).

Cuadro 28. Objetivos y líneas de acción ligados al manejo de recursos naturales en el plan de desarrollo municipal 2015-2018.

Eje Apoyo al campo	
Objetivo	Líneas de acción
Contribuir al aprovechamiento de los recursos agrícolas, acuícolas y ganaderos del municipio, mediante la gestión de recursos y el acompañamiento en el desarrollo de los proyectos que se generen.	<ul style="list-style-type: none"> • Impulsar proyectos productivos para agricultores, ganaderos y acuicultores. • Brindar apoyo y seguimiento a los pequeños productores con técnicos especialistas en cada área. • Promover programas para cultivo de hortalizas para el autoconsumo. • Capacitación para la promoción y comercialización de productos agrícolas y ganaderos de los productores.
Eje Desarrollo económico	
Objetivo	Líneas de acción
Impulsar los diferentes sectores productivos de la economía local y regional.	<ul style="list-style-type: none"> • Promover y fomentar el turismo y el ecoturismo como actividad estratégica. • Apoyar a las pequeñas y medianas empresas que ofrezcan servicios turísticos, con programas de capacitación en el servicio y campañas de promoción turística en coordinación con el Gobierno de Estado.
Eje Servicios públicos	
Objetivo	Líneas de acción
Cuidar el medio ambiente	<ul style="list-style-type: none"> • Programas de reciclaje, PET y separación de basura

En la estructura organizacional, existe una dirección de desarrollo agropecuario, forestal y turismo que, por lo previsto en el reglamento interior de la administración pública municipal, tiene la facultad de formular programas y proyectos en materia agropecuaria, difundir los atractivos turísticos del municipio y concertar proyectos que fortalezcan la preservación de áreas forestales. La dirección estaba conformada

por un director y un coordinador de turismo⁴⁵, ambos asistidos por personal administrativo.

Entre los programas realizados por la dirección se encuentra la dotación de insumos para producir huevo de traspatio, iniciar huertos bio-intensivos y sembrar árboles frutales, los tres con una orientación hacia el autoconsumo. Por otro lado, la dirección tiene la función de recibir solicitudes y documentación de personas interesadas en recibir subsidios provenientes de programas estatales, federales o en donde concurren recursos del municipio y de otro ente público. A través de estos se apoyó principalmente la adquisición de maquinaria, la instalación de infraestructura (depósitos de agua, bodegas, cobertizos) y la compra de animales. Según lo externado en los informes de gobierno 2016 y 2017, el municipio aportó poco más de \$700,000 para la realización de estos programas.

Por su parte, existe una coordinación de ecología dependiente de la dirección de desarrollo urbano y obras públicas. Entre las responsabilidades contenidas en el reglamento municipal están el elaborar e instrumentar programas de educación ambiental, dar seguimiento a programas de reforestación y participar en la creación y administración de zonas de reserva ecológica. La coordinación estaba a cargo de un ciudadano con experiencia previa en organizaciones ambientalistas, sin contar con apoyo adicional. Sin contar con presupuesto para proyectos, las actividades realizadas por el coordinador se enfocaban en promover la separación de residuos sólidos, coordinar campañas de limpieza y atender denuncias por tala de árboles, cacería y contaminación de aguas.

Independientemente de su autonomía, el artículo 134 de la Constitución Política de México establece la obligatoriedad de evaluar, en los tres órdenes de gobierno, los resultados del ejercicio de los recursos públicos (CONEVAL, 2016). En sintonía con lo anterior, documentos institucionales del municipio afirman el seguimiento de los parámetros establecidos por la gestión basada en resultados.

En materia de evaluación, el reglamento interior del municipio estudiado solicita a las distintas áreas desarrollar programas operativos anuales y contar con sistemas de evaluación para lograr mejores resultados en la realización de sus funciones. La instancia interna responsable de dar seguimiento a lo anterior es la contraloría municipal, entidad desconcentrada que supervisa el servicio público para hacer recomendaciones de mejora. En ese sentido, parte de sus funciones es revisar los programas operativos anuales, hacer auditorías de desempeño, e implementar un sistema de evaluación de la gestión que permita verificar, con la participación de las diferentes áreas del municipio, la efectividad de los planes y programas.

⁴⁵ El primero con título de médico veterinario, y el segundo sin título profesional.

3.3.4 Organización de la Sociedad Civil

La Organización de la Sociedad Civil (OSC) que se incorporó en el estudio tiene el objetivo central de conservar la biodiversidad de la reserva de la biosfera al tiempo que promueve el desarrollo económico de la región. Con la finalidad de incrementar su alcance e impacto, la OSC ha liderado la creación de organizaciones adicionales que trabajan de forma paralela pero que se enfocan en nichos específicos, como la diversificación productiva, la conservación *in situ* y la educación ambiental.

La OSC ha resaltado por su capacidad de gestión financiera, contando a lo largo de tres décadas con una larga lista de donantes. En 2016 y 2017, las diferentes organizaciones asociadas a la OSC gestionaron cerca de 9 millones de pesos provenientes de 10 donantes distintos, cuatro de ellos internacionales. A continuación, se enlistan las estrategias de algunos de sus proyectos:

- Compra de predios para convertirlos en reservas privadas destinadas a la conservación *in situ*.
- Conservación basada en el mercado a través de la intermediación para que propietarios de terrenos forestales y ganaderos reciban recursos provenientes de compensaciones de carbono y del pago por servicios ambientales.
- Gestión de recursos y capacitación para micro-operadores turísticos y micro-empresas (cerámica, apicultura, alimentos envasados, herbolaria).
- Gestión de recursos y capacitación para productores agropecuarios y forestales.
- Manejo integral de residuos sólidos

Para la gestión e implementación de los proyectos, la OSC cuenta con un equipo interdisciplinario conformado por profesionistas (administradores, maestros normalistas, ingenieros forestales y agrobiólogos, antropólogos), pobladores locales con años de capacitación y consultores externos.

Debido a que maneja recursos de diversos donantes, desde agencias internacionales hasta dependencias del gobierno mexicano, la OSC interactúa con múltiples enfoques y metodologías para la evaluación externa de resultados. Si bien la norma es el uso de evaluaciones sumativas que verifican la entrega de *outputs* comprometidos, también han tenido evaluaciones para mejorar el desempeño y para medir la obtención de *outcomes*.

De forma interna, la OSC utiliza programas operativos anuales para planificar sus actividades y verificar avances. Además, cuenta con el apoyo de un especialista en monitoreo y evaluación, el cual auxilia a la organización en el diseño de los sistemas de evaluación de múltiples proyectos.

De forma adicional a los requerimientos de sus donantes, la OSC utiliza la metodología del Retorno Social de la Inversión (SROI por sus siglas en inglés) para rendir cuentas sobre los resultados obtenidos. A través de ella, monetiza sus

impactos sociales y ambientales a través de valores sustitutos (*proxy*), informando así cuánto valor se generó a cambio del dinero recibido.

También, debido a que oferta compensaciones de carbono, cuenta con sistemas de evaluación para transparentar la medición del carbono capturado en el suelo y por restauración forestal. Para validar sus protocolos locales, ha contado con la asesoría de diversas instituciones como el Servicio Forestal de los Estados Unidos o la Iniciativa para Transparencia en Acción Climática (ICAT por sus siglas en inglés).

En los siguientes tres capítulos se presentan los resultados de la investigación en un formato de artículo científico. Por ende, en ellos se incluye también una discusión de los hallazgos al contrastarlos con la literatura especializada sobre el tema.

Capítulo cuatro

¿Sirven las evaluaciones para aprender? Influencia de las evaluaciones sobre las decisiones de manejo en un Área Natural Protegida

Are Evaluations Useful for Learning?
The Influence of Evaluations on Management Decisions in a
Natural Protected Area

Eduardo Luna-Sánchez, Margaret Skutsch

Revista: **Sociedad y ambiente**; Año de publicación: **2019**

Liga de descarga:

<https://revistas.ecosur.mx/sociedadyambiente/index.php/sya/article/view/193>

7

**¿Sirven las evaluaciones para aprender?
Influencia de las evaluaciones sobre las decisiones de manejo en un Área Natural
Protegida**

Eduardo Luna Sánchez y Margaret Skutsch

Resumen

El artículo presenta un análisis de la arquitectura de evaluación de la Comisión Nacional de Áreas Naturales Protegidas (CONANP). La finalidad es reconocer factores que promueven e inhiben el uso de evaluaciones para dirigir la toma de decisiones de los guardaparques que administran un Área Natural Protegida (ANP) en el centro de México. Para acceder a la información se utilizaron técnicas de investigación etnográfica, las cuales facilitaron la interacción con el personal de la dependencia para entender el proceso social que influye en la producción y uso de evaluaciones. Se identificó que los guardaparques interactúan con tres tipos de evaluación. Ninguna es utilizada para identificar el impacto de sus acciones. Aun así, su interacción con las evaluaciones les permite identificar aspectos susceptibles de mejora en su manejo del ANP. Algunas barreras a superar se relacionan con el margen de maniobra de los guardaparques para implementar cambios. Por un lado, debido a limitaciones presupuestales que inciden en su capacidad operativa y, por el otro, por las consecuencias negativas de un enfoque evaluativo que exige demostrar resultados cuantificables en el corto plazo.

Palabras clave: áreas protegidas; aprendizaje organizacional; manejo adaptativo; rendición de cuentas; burocracia.

Abstract

The paper presents an analysis of the evaluation system of Mexico's National Commission of Natural Protected Areas (CONANP, by its acronym in Spanish). The purpose is to recognize factors that promote and inhibit the use of evaluations to direct the decision-making of park rangers who manage a Natural Protected Area (ANP, by its acronym in Spanish) in central Mexico. Ethnographic research techniques were used, which facilitated the interaction with the organization's personnel to understand the social process that influences the production and use of evaluations. It was found that park rangers interact with three types of evaluation. Although the evaluation procedures are not used to identify the impact of their actions, this interaction allows them to identify improvement areas in their management of the ANP. Some barriers are related to the margin for maneuver regarding implementation of changes; on the one hand, those related to budgetary limitations that affect their operational capacity and, on the other, to the negative consequences of an evaluative approach that requires demonstrating quantifiable results in the short term.

Keywords: protected areas; organizational learning; adaptive management; accountability; bureaucracy.

Introducción

Las Áreas Naturales Protegidas (ANP) son una de las principales herramientas de conservación en México. Su propósito central es reducir presiones humanas sobre áreas que albergan diversidad biológica y que generan servicios ecosistémicos (Ferraro y Pressey, 2015). Debido a los impactos sociales negativos que ocasiona la exclusión en el acceso a recursos, se ha cuestionado el modelo de “fortalezas de la conservación” que considera a los pobladores locales como una amenaza (Brockington y Wilkie, 2015). En cambio, se ha propuesto un modelo integral de conservación y desarrollo que procura la protección ecológica y el bienestar de los pobladores locales (UNESCO, 2013). Es una tarea compleja cumplir con esa doble expectativa, pues las ANP se ven afectadas por presiones dinámicas que requieren respuestas *ad hoc* (CONANP, 2016).

En este contexto, es necesario orientar las decisiones de manejo de las ANP al reconocer si las acciones están funcionando o no. Esto ha impulsado el interés por las evaluaciones, pues se considera que mejoran la efectividad de las organizaciones al permitir que sus integrantes aprendan y corrijan el rumbo de los proyectos (Chelimsky, 1997; Cousins *et al.*, 2014). Sin embargo, diversas investigaciones sobre el uso de las evaluaciones han cuestionado ese supuesto, pues las evaluaciones con frecuencia son irrelevantes en la toma de decisiones al interior de las organizaciones (Weiss, 1999; Ebrahim, 2005; Guijt, 2008). En general, aprender a partir de las evaluaciones ha sido una buena intención que pocas veces es considerada en términos metodológicos y políticos. Se suele ignorar la forma en que las organizaciones y las relaciones de poder funcionan y, por ello, se sabe poco sobre cómo institucionalizar el aprendizaje derivado de evaluaciones (Guijt, 2008).

A pesar de los llamados para que los practicantes de la conservación mejoren su capacidad de hacer evaluaciones orientadas al aprendizaje, con frecuencia su propósito se ha reducido a la elaboración de reportes de rendición de cuentas (Stem *et al.*, 2005). En algunos escenarios, las consecuencias en caso de obtener resultados negativos, como la pérdida de credibilidad y financiamiento, han sido una barrera para el manejo adaptativo de los proyectos (Craigie *et al.*, 2015; McKinnon *et al.*, 2015).

Este artículo presenta un análisis de la arquitectura de evaluación de la Comisión Nacional de Áreas Naturales Protegidas (CONANP), dependencia responsable del manejo de ANP en México. La finalidad es reconocer factores que promueven e inhiben el uso de evaluaciones para dirigir la toma de decisiones al interior de la dependencia. En particular, se analiza si las evaluaciones influyen en las decisiones de manejo de los guardaparques que administran un ANP en el centro de México. La visión de los guardaparques es fundamental, pues deberían ser los principales usuarios de evaluaciones que sirvan para adaptar el manejo en contextos que cambian constantemente. Sin embargo, su perspectiva ha sido poco estudiada.

En México, la política de evaluación del gobierno federal considera que los usuarios de las evaluaciones son funcionarios de gobierno de primer nivel (Cardozo, 2015). En torno de la CONANP, las investigaciones se han centrado en esa escala, al estudiar las evaluaciones externas de sus programas de subsidio (Orta *et al.*, 2013; Sotelo *et al.*, 2015). En ese sentido, no hay investigaciones sobre el sistema de evaluación diseñado por la dependencia. Por estos motivos, es relevante analizar el sistema de evaluación de la CONANP desde el punto de

vista de los guardaparques. Es importante enfatizar que la investigación no pretende evaluar la efectividad de los proyectos de la CONANP en el ANP estudiada. En cambio, se exploran las causas por las que personal de la dependencia utiliza, o no, evaluaciones como una herramienta de aprendizaje. Las preguntas que se busca responder son: ¿Cómo se hacen y usan evaluaciones relacionadas con el manejo del ANP?, ¿En qué grado las evaluaciones se han utilizado para el manejo adaptativo del ANP?, ¿Cuáles son las principales barreras para que los guardaparques aprendan a partir de esas evaluaciones?

Metodología

Los estudios de caso son un tipo de investigación cualitativa que estudia un fenómeno en su propio contexto. El método permite vincular datos empíricos con proposiciones teóricas. Por ello, es útil para verificar si las teorías funcionan en el mundo real (Yin, 2009). En el marco de un estudio de caso, se utilizaron técnicas de investigación etnográfica para conocer la influencia que tiene el sistema de evaluación de la CONANP sobre la toma de decisiones de los guardaparques que administran un ANP en el centro de México.

Las técnicas de investigación (entrevistas semiestructuradas, observación participante y revisión de documentos) facilitaron la interacción con el personal de la dependencia para entender cómo su comportamiento es afectado por reglas, procedimientos, valores, creencias, y relaciones de poder (Lewis *et al.*, 2003; Mosse, 2004). Las entrevistas se hicieron a personal directivo y operativo que labora o laboró en la CONANP. El criterio de selección de los entrevistados fue la relación con diferentes procedimientos de evaluación. Sin embargo, la selección incluyó personal que se encuentra en diferentes posiciones¹, dada la posible existencia de diferencias intraorganizacionales. Las entrevistas permitieron recolectar información sobre la puesta en práctica de las evaluaciones, entremezclada con la opinión y reflexiones de los entrevistados en torno al tema.

A través de la observación participante, fue posible explorar la cotidianidad e interacciones del personal de la dependencia. Esto permitió hacer descripciones detalladas de diferentes aspectos relacionados con la cultura organizacional. La recopilación de documentos relacionados con las evaluaciones se llevó a cabo a través de una exploración de la página web institucional de la CONANP, y a través de solicitudes directas al personal directivo de la dependencia. De esa manera se tuvo acceso a evaluaciones externas, informes, y documentos de planeación estratégica.

A partir de una revisión de literatura, se seleccionaron variables relacionadas con el uso de las evaluaciones. Su estudio es complejo y contextual; sin embargo, se ha destacado la relevancia de distintos factores metodológicos, organizativos e institucionales que se entrelazan entre sí (Cousins y Leithwood, 1986; Shulha y Cousins, 1997; Cardozo, 2015). Sobre cada uno de ellos se desarrollaron ítems para estructurar las observaciones, las entrevistas y la revisión de documentos. Las fuentes de información fueron las siguientes:

¹ A nivel de ANP se entrevistó al director y a técnicos del área. A nivel directivo se entrevistó a personal de la Dirección de Evaluación y Seguimiento de la CONANP.

- Once entrevistas semiestructuradas al personal de la CONANP (personal directivo que labora o laboró en la dependencia, y guardaparques del ANP elegida).
- La revisión de documentos institucionales (CONEVAL y CONANP).
- Observación participante al acompañar a los guardaparques a salidas de campo para implementar programas de subsidio.

El acceso a la información² fue el principal reto de la investigación. Las evaluaciones suelen generar ansiedad, pues reflejan una preocupación por el estatus (Donaldson *et al.*, 2002). A nivel organizacional, esto se traduce en el temor de perder credibilidad y de recibir sanciones en caso de no cumplir con expectativas de desempeño (McKinnon *et al.*, 2015). Para que los integrantes de una organización puedan involucrarse en la reflexión sobre su práctica, es necesario que existan altos niveles de confianza, de forma que sea seguro compartir abiertamente su opinión y experiencia (Serrat, 2010; Kusters *et al.*, 2011). Por ello, se tomó la decisión de garantizar el anonimato de los guardaparques que participaron en la investigación. Como consecuencia, no se brinda información específica sobre el ANP que manejan. Las limitaciones del enfoque utilizado son las propias de cualquier investigación exploratoria de un problema complejo. No se puede esperar como resultado la identificación de relaciones causales entre las variables, pero sí la detección de patrones que permitan cuestionar supuestos existentes y generar nuevas hipótesis de trabajo.

El sistema de evaluación de la CONANP

La CONANP es un órgano descentralizado de la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). La dependencia tiene a su cargo la conservación de la biodiversidad en los ecosistemas más representativos de México. Actualmente, a través de nueve direcciones regionales, la CONANP administra 182 ANP de carácter federal que abarcan más de 90 millones de hectáreas (CONANP-GIZ, 2018). A escala local, las direcciones de ANP son el brazo operativo de la dependencia. En ellas laboran guardaparques que implementan acciones de conservación, idealmente, siguiendo un programa de manejo orientado a la atención de problemáticas locales. Uno de los principales retos de la dependencia es cubrir la brecha financiera para lograr el manejo efectivo de las ANP. A pesar de que la superficie protegida continúa creciendo, en 2017 el presupuesto de la CONANP tuvo una reducción del 22% (Chacón y Garduño, 2017). En este contexto de limitación presupuestal, las evaluaciones tienen un papel estratégico para orientar las decisiones de manejo e incrementar la efectividad de su protección.

En diferentes documentos institucionales es claro que la CONANP considera a la evaluación como una herramienta valiosa para poder cumplir con su misión (CONANP, 2010a, 2010b, 2014a, 2016). La dependencia afirma que la evaluación es una herramienta de aprendizaje organizacional, la cual permite tomar decisiones con un enfoque adaptativo, mejorando el manejo ante ambientes cambiantes (CONANP, 2014a). Además, desde un enfoque de gestión

² El acceso a la información se comprende como la suma de la disponibilidad de información y la voluntad para compartirla en las organizaciones.

basada en resultados, la evaluación permite rendir cuentas sobre los resultados obtenidos, lo que asegura fuentes de financiamiento presentes y facilita la gestión de recursos adicionales (CONANP, 2010a).

En la estructura operacional de la dependencia, es atribución de la Dirección de Evaluación y Seguimiento (DES) proponer los procedimientos para hacer evaluaciones. Esta instancia es fundamental, pues articula todo el sistema de evaluación de la CONANP. El sistema de evaluación tiene distintos niveles de análisis (nacional o ANP) y propósitos (planeación estratégica, manejo adaptativo, rendición de cuentas, evaluación del desempeño). Esto incide en la metodología de las evaluaciones y define quién es su usuario potencial, ya sea personal directivo o guardaparques (CONANP, 2010a). En ese sentido, para valorar de forma adecuada los efectos del sistema de evaluación sobre la toma de decisiones de los guardaparques, es importante tener clara la finalidad de cada tipo de evaluación (Cuadro 1).

Cuadro 1. Tipos de evaluación de la CONANP.

Tipo de evaluación	Escala	Propósito	Usuario potencial
Evaluaciones externas a programas de subsidio	Nacional	Planeación estratégica Rendición de cuentas	Personal directivo
Evaluación de programas de manejo de ANP	ANP	Manejo adaptativo	Guardaparques
Evaluación del Programa Operativo Anual (POA)	Nacional ANP	Manejo adaptativo Rendición de cuentas Evaluación del desempeño Planeación estratégica	Personal directivo Guardaparques
Evaluación de la efectividad en el manejo	Nacional ANP	Manejo adaptativo Planeación estratégica Rendición de cuentas	Personal directivo Guardaparques

Fuente: Elaboración propia.

Evaluaciones externas a programas de subsidio

Las evaluaciones externas a programas de subsidio forman parte de la política de evaluación del gobierno federal. Su propósito es evaluar la obtención global de resultados y emitir recomendaciones para mejorar el desempeño de los programas (Cejudo, 2011). Por su escala y propósito, el usuario potencial es el personal directivo de la dependencia, quien puede retomar la información para rediseñar los programas. En la CONANP, es responsabilidad de la DES atender las recomendaciones de mejora, con la finalidad de integrarlas a la planeación estratégica de la dependencia.

Si bien, las evaluaciones externas no se dirigen hacia los guardaparques, sí permiten identificar características del sistema de evaluación de la dependencia que inciden en su disponibilidad de información para tomar decisiones.

El programa de subsidio más importante de la CONANP, en términos presupuestales, es el Programa de Conservación para el Desarrollo (PROCOCODES). La Auditoría Superior de la Federación ha hecho dos auditorías de desempeño (2012 y 2016) y el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) ha efectuado diversos tipos de evaluación a lo largo de nueve años³. Las evaluaciones han observado, en repetidas ocasiones, que el indicador de propósito utilizado⁴ no permite medir los efectos generados. En este sentido, las recomendaciones se han dirigido hacia dos aspectos: 1) modificar el diseño del programa, de forma que el indicador de resultado refleje el efecto causado por la intervención, 2) llevar a cabo una evaluación de impacto que mida los cambios atribuibles al programa.

Evaluación de programas de manejo de ANP

El reglamento de la Ley General del Equilibrio Ecológico y la Protección al Ambiente (LGEEPA) en materia de ANP indica que el programa de manejo (PM) es el instrumento rector de planeación y regulación de las ANP. Es responsabilidad de las direcciones de ANP elaborar, evaluar y actualizar el PM de su área, pues en él se basan su planeación operativa (CONANP, 2014b). El reglamento indica que el PM será revisado por lo menos cada cinco años con el objeto de evaluar su efectividad y proponer posibles modificaciones. Para cumplir con este mandato, la CONANP desarrolló la cédula de evaluación de la efectividad del PM. En ella, los guardaparques evalúan el cumplimiento de los objetivos trazados en los diferentes componentes del programa y, en caso de ser necesario, explican las razones por las que hay que replantear las estrategias y acciones.

Para hacer el análisis, es necesario que los PM cuenten con un sistema de “seguimiento y monitoreo de los proyectos” que permita evaluar los avances anualmente (CONANP, s/f). Es atribución de las direcciones conformar un sistema de información que concentre datos sociales, económicos, y biológicos del ANP que administran, de forma que pueda utilizarse la información para evaluar la efectividad del manejo. Además, las direcciones deben promover el establecimiento de un consejo asesor, el cual colabore en el seguimiento del PM del área (CONANP, 2014b). Por su escala y propósito, el principal usuario de este tipo de evaluaciones es el guardaparques, pues brinda insumos para identificar áreas de mejora en el manejo que hace del ANP.

³ La Ley General de Desarrollo Social (2004) establece que el CONEVAL tiene la atribución de evaluar los programas sociales. Por su parte, la Ley de Fiscalización y Rendición de Cuentas de la Federación (2009), establece que la ASF evaluará el desempeño de los programas federales.

⁴ Número de localidades que reciben apoyos del programa en relación al número total de localidades asentadas en ANP.

Evaluación del Programa Operativo Anual (POA)

Como procedimiento organizacional, cada dirección de ANP tiene la obligación de elaborar un Programa Operativo Anual (POA) que se base en los objetivos de su PM. El instrumento permite ajustar las acciones al presupuesto anual disponible para el ANP. El POA tiene dos propósitos adicionales:

Alinear las acciones de las direcciones de ANP al Programa Nacional de ANP (PNANP). La DES tiene la responsabilidad de comunicar el grado de cumplimiento de las metas establecidas en el PNANP, documento de planeación estratégica de la dependencia. Por ello, para facilitar la integración de la información, desarrolló el Sistema General de Programas Operativos Anuales (SGPOA) (Cuadro 2). A través de un catálogo de actividades y unidades de medida, el sistema estandariza la información que las direcciones de ANP pueden reportar, con la finalidad de que sirva para alimentar los indicadores de resultado con los que se mide el desempeño global de la dependencia.

Evaluar el desempeño de las direcciones de ANP. La DES y las direcciones regionales utilizan el grado de avance de las actividades comprometidas en el SGPOA como un mecanismo para evaluar trimestralmente el desempeño de las direcciones de ANP.

Cuadro 2. Algunos indicadores del Programa Nacional de Áreas Naturales Protegidas (PNANP).

Objetivo	Indicador
Conservación y manejo de la biodiversidad	Programas de Acción para la Conservación de Especies (PACE) elaborados
	ANP con acciones de restauración
	ANP que implementan monitoreo de especies prioritarias
	Comités de vigilancia comunitaria para la protección y monitoreo de la biodiversidad
Economía de la conservación	Proyectos productivos sustentables apoyados con Programas de subsidio
Comunicación, educación, cultura y participación social para la conservación	Porcentaje de ANP con Programa de educación ambiental
	Consejos Asesores y otros mecanismos de gobernanza formalmente instalados

Fuente: CONANP, 2017a

Como herramienta de visualización, el sistema genera un puntaje que clasifica las áreas protegidas en tres rangos (rojo, amarillo y verde), el cual representa el porcentaje de cumplimiento de las actividades. Esto permite visualizar con facilidad retrasos y acciones no

implementadas. En su caso, los guardaparques pueden explicar las razones por las que es necesario reprogramar o cancelar las actividades definidas en su POA. Anualmente, las evaluaciones de desempeño se utilizan para retroalimentar a los guardaparques y tomar decisiones presupuestarias. En ese sentido, en caso de que una dirección de ANP haya planteado una serie de metas no cumplidas, esa información puede utilizarse como evidencia para asignar un presupuesto menor, con la finalidad de ajustarlo a sus capacidades de implementación. Del mismo modo, en caso de cumplir con todas las metas, es posible que la dirección de ANP reciba un aumento presupuestal (Personal directivo, comunicación personal, 8 de febrero de 2017).

Los usuarios potenciales de la evaluación del POA son tanto los guardaparques como el personal directivo de la dependencia. En la escala local, los guardaparques podrían retomar la información para evaluar el cumplimiento de los objetivos de su programa de manejo y retroalimentar la planeación operativa. A escala nacional, el personal directivo la utiliza para rendir cuentas sobre los resultados de la dependencia, para evaluar el desempeño de su personal, y, posiblemente, como un insumo para su proceso de planeación estratégica.

Evaluación de la efectividad en el manejo

México adquirió el compromiso de cumplir con la meta 11 de Aichi ante el Convenio de Diversidad Biológica. Esta meta establece que para el año 2020, el 17% de la parte terrestre de México, así como el 10% de su superficie marina estarán bajo un régimen de conservación, y que esa superficie será administrada de manera efectiva. Por ello, uno de los indicadores estratégicos del PNANP es el número de ANP incorporadas a un proceso de evaluación de la efectividad (CONANP, 2014c). Para la CONANP, la evaluación de la efectividad de manejo sirve para estimar el progreso y apoyar a los directores en el manejo adaptativo de las ANP (CONANP, s.f.). A lo largo de los años la dependencia ha utilizado diferentes metodologías, entre ellas la ficha de evaluación de condiciones ecológicas (Scorecard), el Índice de Implementación y Gestión (Indimapa), o la Herramienta de Seguimiento de la Efectividad del Manejo (METT) (Cuadro 3).

Por otro lado, como indicador de la efectividad en el manejo, la CONANP implementa programas de monitoreo biológico articulados a los POA de las ANP. En ese sentido, el seguimiento de la población de especies en riesgo o de especies indicadoras de las condiciones del hábitat permite establecer si las acciones de conservación han tenido un efecto o no. Otro indicador utilizado por la CONANP es la tasa de cambio de las cubiertas del suelo en las ANP (CONANP, 2014c).

Para contar con un estándar propio, la DES diseñó el sistema permanente de evaluación de la efectividad en el manejo de las Áreas Naturales Protegidas Federales de México (i-efectividad). El sistema cuenta con 48 indicadores que se retomaron de otros instrumentos de evaluación⁵, con ellos se calcula un índice que clasifica la efectividad en cinco rangos (CONANP, 2017b). El proyecto se encuentra a prueba con áreas piloto, pero la intención es que el método se aplique cada tres años en las ANP administradas por la CONANP (Personal directivo, comunicación personal, 26 de junio de 2017).

⁵ Treinta de ellos de la metodología METT.

Cuadro 3. Metodologías de evaluación de la efectividad que han sido utilizadas por la CONANP.

Metodología	Descripción	Referencia
Scorecard	Busca evaluar las condiciones y tendencias de tres elementos en las ANP: agua, hábitat y recursos biológicos.	Comisión para la Cooperación Ambiental, 2011
Indimapa	Su objetivo es establecer si las ANP terrestres de América Latina poseen las condiciones institucionales, normativas y operativas necesarias para ser efectivas. Para ello, obtiene información sobre distintos indicadores vinculados a la gestión de un ANP ⁶ . La herramienta genera un índice que clasifica las áreas protegidas en tres rangos (rojo, amarillo y verde) correspondientes a un nivel bajo, mediano y alto de implementación y gestión.	Organización Latinoamericana y del Caribe de Entidades Fiscalizadoras Superiores, 2015
METT	Evalúa la efectividad de manejo de un ANP al revisar la presencia de diferentes aspectos que son condiciones necesarias de un buen ciclo de gestión de proyectos de conservación. Esto permite identificar acciones que rectifiquen debilidades en el manejo.	Stolton y Dudley, 2016
Evaluación de áreas marinas protegidas	Contempla indicadores biofísicos, socioeconómicos y de gobernabilidad para evaluar la efectividad de áreas protegidas marinas.	Pomeroy <i>et al.</i> , 2004

Fuente: Elaboración propia.

El usuario potencial de este tipo de evaluaciones es tanto el personal directivo como los guardaparques. A escala local, el sistema permite identificar fortalezas y áreas de oportunidad en la gestión de las áreas. A escala nacional, la metodología pretende ser un medio de verificación de la meta 11 de Aichi y, a la vez, ser útil para elaborar reportes de rendición de cuentas sobre los resultados de la dependencia (CONANP, 2017b).

Caracterización del equipo de guardaparques

La dirección de ANP estudiada forma parte de la Dirección Regional Centro y Eje Neovolcánico, que agrupa las áreas de 11 estados del centro de la República Mexicana. Se

⁶ Entre los que se encuentran: Plan de manejo, Recursos humanos, Recursos financieros, Estructura administrativa, y Monitoreo de la biodiversidad.

conforma por diez guardaparques, los cuales cuentan con un perfil profesional relacionado con el manejo de recursos naturales (biología, ingeniería forestal, veterinaria) y, en menor medida, con áreas administrativas (administración, turismo).

En la estructura organizacional del ANP existe una dirección, una subdirección, una unidad administrativa y una jefatura de proyectos. Estas subunidades son responsables de la coordinación, planificación y supervisión de las actividades del núcleo operativo. El área operativa se compone de seis técnicos, cada uno con diferentes funciones: educación ambiental, monitoreo biológico, revisión de evaluaciones de impacto ambiental, regulación de actividades turísticas, promoción y ejecución de programas de subsidio⁷, vigilancia, protección contra incendios, y restauración de ecosistemas. Sin embargo, debido a la extensión del ANP⁸ y a restricciones presupuestales, el personal tiene una carga de trabajo que le lleva a cubrir múltiples funciones a la vez.

El equipo de guardaparques cuenta con experiencia en la planificación de proyectos utilizando el marco lógico, que es una manera de presentar la lógica causal de objetivos, actividades y resultados, utilizando indicadores para medir los avances. Además, ha recibido más de 100 horas de capacitación en gestión y evaluación de proyectos. En específico, el personal directivo ha asistido a talleres sobre el modelo de gestión Capacity Works⁹, en donde ha participado en la revisión y elaboración de POA e indicadores de impacto.

Resultados

TIPOS DE EVALUACIÓN

Los guardaparques del ANP interactúan con tres tipos de evaluación:

1. Evaluación del Programa Operativo Anual (POA). La elaboración del POA es un procedimiento organizacional obligatorio, con una frecuencia anual de evaluación. Se utiliza como insumo para retroalimentar la planeación operativa de los guardaparques y, a la vez, como un mecanismo para evaluar el desempeño global de la dirección de ANP.
2. Evaluación del PM. De 2014 a la fecha el equipo de guardaparques trabaja en la actualización del PM, el cual tiene más de 15 años de antigüedad. Como parte del procedimiento elaboraron la cédula de evaluación de la efectividad del PM.
3. Evaluación de la efectividad del manejo.
 - Indimapa. Se ha llevado a cabo anualmente entre 2014 y 2017, con base en las respuestas del director del ANP a cuestionarios de opción múltiple. En este periodo la dirección de ANP se ha ubicado en un rango alto de implementación y gestión.

⁷ Entre los que se encuentra el PROCODES, el Programa de recuperación de especies en riesgo y el Programa de Conservación de Maíz Criollo.

⁸ El ANP cuenta con cerca de 25 000 hectáreas de zona núcleo.

⁹ Metodología de planeación estratégica de la Agencia de Cooperación Alemana al Desarrollo (GIZ, por sus siglas en Alemán).

- i-efectividad. La evaluación estaba en proceso al momento de la investigación. La metodología solicita responder un cuestionario de opción múltiple para valorar la efectividad del 2013 al 2016. La evaluación se enviará a la DES para recibir su retroalimentación sobre fortalezas y áreas de mejora.

FACTORES QUE INCIDEN EN EL USO DE EVALUACIONES

A continuación, se presentan una serie de factores que inciden en el uso de las evaluaciones para retroalimentar la planeación operativa de los guardaparques (Cuadro 4).

Cuadro 4. Factores que inciden en el uso de las evaluaciones en el ANP

Tipo de evaluación	Propósito	Barreras		Efectos de la evaluación
		Diseño de las evaluaciones	Capacidad organizacional	
<i>Evaluación del POA</i>	Planeación operativa Evaluación del desempeño	Indicadores	Seguimiento	(+) Aprendizaje (-) Comportamiento administrativo estratégico
<i>Evaluación de programas de manejo de ANP</i>	Actualización del PM		Carga de trabajo	(-) Visión de corto plazo (-) Atención de la agenda local
<i>Evaluación de la efectividad en el manejo</i>	Identificar fortalezas y áreas de mejora en la gestión del ANP	Obtención de resultados	Recursos	(+) Aprendizaje
			Manejo de información y conocimiento	(?) Hace falta un análisis futuro de la influencia de este tipo de evaluación

Fuente: elaboración propia.

- *Diseño de las evaluaciones*

En este apartado se describen aspectos vinculados con la adecuación metodológica de las evaluaciones. Este aspecto busca identificar si su diseño permite generar información sobre

la efectividad de las acciones, con la finalidad de que los guardaparques utilicen ese conocimiento para hacer ajustes.

Indicadores de gestión. Las unidades de medida del POA sirven para medir la obtención de resultados inmediatos (*outputs*), pero no para medir el efecto de las acciones (*outcomes*) (Cuadro 5).

Cuadro 5. Actividades del Programa Operativo Anual (POA).

Línea estratégica	Objetivo	Actividad	Unidad de medida	Medio verificación
Vigilancia	Prevenir y disminuir la ocurrencia de los ilícitos ambientales en el Área Natural Protegida	Gestionar la integración de comités de vigilancia comunitarios	Número de comités integrados	Convenios
Restauración de ecosistemas	Promover la recuperación de ecosistemas originales	Apoyar la producción de plantas nativas	Número de viveros operando	Convenio de Concertación, memoria fotográfica
Recuperación de especies en riesgo	Contribuir en la conservación de especies clave	Realizar actividades de conocimiento (monitoreo) y educación ambiental (talleres)	Número de acciones	Reporte y memoria fotográfica

Fuente: Extracto del POA 2017.

Estos indicadores, estandarizados y evaluados por el SGPOA, no son útiles para estimar la efectividad del manejo. La limitación de este tipo de evaluación es reconocida por los guardaparques:

“... al estar evaluando cuánto trabajo se hizo en número de reuniones, de talleres, de hectáreas cercadas, se está evaluando el trabajo y no los resultados” (Guardaparques #1, comunicación personal, 14 de enero de 2017).

“... los indicadores que se usan no son indicadores de conservación, pues una actividad de educación ambiental en la que participaron cien personas, no quiere decir que esas cien personas ya fueron concientizadas. A lo mejor el indicador adecuado sería que estas personas están haciendo acciones de conservación, pero obviamente eso te implicaría un mayor esfuerzo de medición” (Guardaparques #2, comunicación personal, 12 de enero de 2017).

La carencia de información sobre los efectos de las acciones es una limitante para evaluar la efectividad del PM. Como resultado, la evaluación elaborada por los guardaparques se

concentró en el cálculo de porcentajes de implementación al comparar las acciones realizadas con las acciones originalmente previstas en el PM. De cualquier forma, el procedimiento de la cédula de evaluación no especifica la forma en que se medirá la “efectividad” o el “cumplimiento” de los objetivos. Tampoco el PM del ANP ofrece orientación en ese sentido, pues la mayor parte de sus indicadores sólo estiman la obtención de resultados inmediatos (p.e. número total de acciones aplicadas derivadas del PM, número de programas de conservación, número total de investigaciones científicas básicas y aplicadas, o el número total de proyectos productivos sustentables).

El uso exclusivo de indicadores de gestión ha sido reconocido como una limitante por la dependencia. Una revisión de los indicadores del PNANP 2007-2012 concluyó que era necesario reformular varios indicadores para que pudieran medir la efectividad de las acciones (CONANP, 2010b). La Agencia de Cooperación Alemana para el Desarrollo (GIZ, por sus siglas en Alemán) elaboró una propuesta para fortalecer el SGPOA con la integración de indicadores de “manejo adaptativo” y la georreferenciación de actividades (CONANP-GIZ, 2018). Actualmente se encuentra en fase piloto en otras direcciones de ANP.

Obtención de resultados. Por su diseño, las metodologías para evaluar la efectividad empleadas por los guardaparques (INDIMAPA; i-efectividad) no permiten reconocer si las acciones generan los resultados de conservación esperados. En cambio, las herramientas identifican la presencia de elementos que se consideran imprescindibles en un buen ciclo de gestión de proyectos de conservación (por ejemplo: existencia de mecanismos de seguimiento que retroalimenten las decisiones de los guardaparques, suficiencia de recursos para la operación, contar con un PM). Por ello, en sentido estricto no valoran la efectividad de las acciones, sino precondiciones organizacionales que son la base para obtener resultados. Su propósito es identificar áreas de oportunidad para mejorar el manejo que hace la dirección de ANP en el área.

- *Capacidad organizacional*

Se identificaron una serie de factores relacionados con la capacidad de la dependencia para brindar medios y oportunidades para que las evaluaciones ocurran (Serrat, 2010). Esto incide sobre la disponibilidad de información para que los guardaparques tomen decisiones.

Falta de procedimientos de seguimiento. No existe un procedimiento organizacional para dar seguimiento a las acciones. Uno de los factores que influye en esa limitación es el año fiscal de la administración pública.

“En cuestiones de seguimiento nosotros básicamente no hacemos nada, aplicamos los programas y terminando el año fiscal y el papeleo, básicamente no volvemos a ir a las comunidades, a ver cómo funcionó, quién si lo sigue haciendo, o quién tiene problemas” (Guardaparques #3, comunicación personal, 7 de marzo de 2017).

“... muchas veces por la dinámica que traemos en los propios programas, y en general en la administración pública, no nos damos a estas tareas fundamentales de decir -está bien, aplicamos el programa, aplicamos tanto recurso, tantos proyectos, ahora ¿cuál es el impacto real que se observa de la aplicación en cada uno de estos proyectos?” (Guardaparques #4, comunicación personal, 12 de febrero de 2018).

Carga de trabajo. La saturación laboral de los guardaparques es uno de los motivos que dificulta el seguimiento de las acciones. A diferencia de la Comisión Nacional Forestal (CONAFOR), que se apoya en la contratación de asesores técnicos externos para diseñar e implementar las acciones de sus programas de subsidio, en la CONANP son los guardaparques los que absorben esa carga. Esto implica que participan en el diagnóstico, diseño, acompañamiento y supervisión de los proyectos, comprometiendo su disponibilidad de tiempo.

“En cada uno de los técnicos no recae una sola función, hacemos dobles o triples funciones porque somos pocos. El área es muy extensa y el territorio es poco accesible en algunos puntos” (Guardaparques #4, comunicación personal, 12 de febrero de 2018).

Falta de recursos. Uno de los motivos transversales por los que no se pueden llevar a cabo actividades de seguimiento es la falta de recursos. Por ello, en el ANP el presupuesto se destina para cubrir gastos administrativos y acciones de manejo.

“Sería lo ideal evaluar si realmente las áreas están cumpliendo con sus objetivos, pero ahora el presupuesto está súper bajo, no hay dinero para salir a campo, las pocas salidas son muy limitadas, porque ya no hay recursos, no hay gasolina, no hay viáticos. Ahora, agregarle que se tuviera que estar revisando en campo si se tiene impacto, se me hace muy complicado” (Guardaparques #2, comunicación personal, 12 de enero de 2017).

“... Véelo como para ti, si tienes poco dinero para alimentarte ¿Destinarías la mitad en una evaluación psicológica para ver cómo andas? Yo creo que no” (Personal directivo, comunicación personal, 26 de junio de 2017).

Manejo de información y conocimiento. La dirección de ANP no cuenta con un sistema para almacenar y distribuir la información y conocimiento generados con la implementación de las acciones. Esto incide sobre la cantidad de información disponible para la toma de decisiones futura. El conocimiento se refleja en la experiencia de los guardaparques, pero está vulnerable en caso de que éstos dejen la dependencia.

“... no hay alguna plataforma donde se comparta la información que se genera, simplemente, al estar todos dentro de la oficina, se hacen algunas reuniones para ver cómo vamos... al final muchas cosas se determinan por lo que uno sabe aunque no esté escrito, es muy difícil que un técnico realmente aterrice un informe físico en el que explique lo que sabe” (Guardaparques #3, comunicación personal, 7 de marzo de 2017).

“Si uno busca qué ha sucedido en administraciones pasadas pues no tenemos información, la información muchas veces se va con las personas, a veces no queda todo documentado...” (Guardaparques #4, comunicación personal, 12 de febrero de 2018).

Esto es relevante, pues en los últimos nueve años, la Dirección de ANP ha tenido cuatro directores diferentes, reflejando cierta inestabilidad en el puesto. Además, en el periodo del director actual, ha tenido lugar rotación de personal, principalmente ocasionada por las bajas percepciones laborales del personal operativo.

EFFECTOS DE LAS EVALUACIONES SOBRE LA TOMA DE DECISIONES DE LOS GUARDAPARQUES

En este apartado se describen los efectos de las evaluaciones sobre el comportamiento de los guardaparques. En específico, se rastrea la influencia de los procedimientos sobre sus decisiones de manejo.

- *Aprendizaje*

A falta de actividades de seguimiento, la evaluación interna del POA permite a los guardaparques recuperar el aprendizaje acumulado a través de la implementación de los proyectos. La reflexión colectiva sobre las experiencias del año anterior genera información útil que les permite hacer ajustes en su POA, con la finalidad de incrementar su efectividad.

“El POA se genera entre todo el equipo de trabajo. Se hizo un ejercicio diagnóstico de cuáles son las áreas que cubrimos, cuáles no hemos podido alcanzar, cuáles son los problemas que se presentan y las estrategias para poder cubrirlos. Entonces cada uno en su área plantea sus necesidades de equipo, de capacitación o de recursos para poder realizar las actividades, y entre todo el equipo vamos dando opiniones sobre cómo sería la mejor manera de que impacte” (Guardaparques #5, comunicación personal, 10 de enero de 2018).

A través de sesiones de trabajo donde se analiza lo que salió bien, lo que falló, y lo que podría mejorar, los guardaparques aprenden a mejorar el manejo de los proyectos, haciendo adaptaciones en su práctica.

“... nos vamos hacia atrás y vemos que quienes aplicamos el recurso, quienes hicieron las cuestiones de dar talleres y dar el equipamiento fue mal dirigido y fue mal hecho, por eso a los apicultores aquí les cuesta mucho subir y producir, han tenido pocas herramientas realmente eficientes. Entonces, todo ese recurso debería hablar por sí solo a la hora que tú muestres a tus apicultores exitosos, pero nos dimos cuenta de que no había sido así” (Guardaparques #3, comunicación personal, 7 de marzo de 2017).

De esta manera, si bien no hay un seguimiento formal, canalizado a través de una evaluación, esto no quiere decir que los guardaparques no tengan conocimiento sobre el funcionamiento de los proyectos, y sobre las variables que intervienen en su éxito o fracaso. Los momentos de intercambio reflexivo suceden a través de reuniones en la oficina, fomentadas por la necesidad de generar el siguiente POA, pero también ocurren al compartir un vehículo, o una comida. Por su parte, al percatarse de que no contaban con información para evaluar algunos de los componentes del PM, la evaluación de la efectividad del PM permitió a los guardaparques identificar deficiencias en los procedimientos para almacenar información. Además, independientemente de las limitaciones de la evaluación, el proceso sirve a los guardaparques como un medio para reflexionar en las estrategias que requiere la actualización del PM.

- *Comportamiento administrativo estratégico*

Un efecto difícil de detectar, pero que fue mencionado por los guardaparques, es la modificación del comportamiento para evitar sanciones por incumplir con expectativas de

desempeño. Específicamente, como respuesta al uso de la evaluación del POA para juzgar la capacidad de las direcciones de ANP y tomar decisiones presupuestales.

“Cuando la calificación de tu POA te aseguraba un mejor presupuesto para el año entrante, ahí entonces empezó la gente a alimentar el sistema para tener una mejor calificación” (Personal directivo, comunicación personal, 8 de febrero de 2017).

Esto puede verse reflejado en el diseño del POA, al privilegiar la selección de actividades “seguras”, sin problemas de implementación. Esta decisión permite minimizar el riesgo de no cumplir, pero puede orientar las acciones hacia resultados superficiales y efímeros.

“Lo que sí uno cuidaba mucho era que lo que está en tus manos, sí asegúralo, que nada más dependa de ti como área. Por ejemplo, —voy a gestionar a través de la CONAFOR o de SAGARPA tantas cosas—, pero sabes que eso no está en tus manos, entonces usaba palabras como gestionar. Gestionar quiere decir, —yo gestioné pero no se logró, nada más se metió el oficio—... Ya cuando dices ejecutar, quiere decir que sí lo hiciste o lo vas hacer... con el paso del tiempo vas aprendiendo a que no te ahorquen” (Guardaparques #2, comunicación personal, 12 de enero de 2017).

- *Visión de corto plazo*

Al vincular la evaluación de desempeño con la comprobación de actividades, algunos guardaparques identificaron el riesgo de privilegiar resultados inmediatos, perdiendo de vista los objetivos de largo plazo estipulados en el PM.

“Yo siento que sí hay un poco la tendencia a cumplir con los números establecidos... a lo mejor yo sí cumplí con aplicar los cinco proyectos que me propuse, pero bueno el siguiente año qué va a suceder con estos proyectos, ¿siguen trabajando?, ¿realmente están dando las ganancias que esperábamos que tuviera la gente?, ¿son proyectos que tendríamos que seguir apoyando?” (Guardaparques #4, comunicación personal, 12 de febrero de 2018).

- *Atención de la agenda local*

Debido a un interés por cierto tipo de acciones e indicadores, algunas actividades del POA son definidas por otros niveles de la estructura organizacional y no por el equipo de guardaparques.

“Por ejemplo, acá yo quisiera hacer monitoreo de cuestiones estratégicas que no son especies, yo me he metido mucho a la cuestión del agua pero a nivel regional mi estudio no es bien visto, pues no son especies que puedan cacaraquear los entes arriba. Si yo no alimento buenos indicadores para la CONANP en los que estén obteniendo buenos resultados y se vea reflejado que se están haciendo acciones para la conservación pues entonces no les sirven mucho los esfuerzos que tú haces. Si pones la actividad [en el POA], y no les hace sentido, entonces no le destinan recursos a lo que estás haciendo y tienes que buscar por fuera” (Guardaparques #3, comunicación personal, 7 de marzo de 2017).

“... vienen directrices de arriba... ahora fue ‘tienen que hacer proyectos productivos’ y hemos estado haciendo más proyectos productivos. Algunos a veces no tienen fundamento, uno no

les ve mucha seguridad, pero por cuestiones de arriba y porque son los que nos llegaron, pues tenemos que ver cómo los apoyamos” (Guardaparques #4, comunicación personal, 12 de enero de 2017).

Discusión

Un factor que incide en el uso de las evaluaciones es la pertinencia de su metodología para conseguir los fines deseados (Chelimsky, 1997). En ese sentido, si se busca el manejo adaptativo, se requiere que las evaluaciones identifiquen relaciones de causalidad para ajustar la dirección de los proyectos. Por su diseño, la evaluación del POA se centra en la valoración del desempeño de los guardaparques. Las métricas utilizadas miden avances en la implementación al aportar información descriptiva que documenta la generación de productos comprometidos (por ejemplo: talleres de capacitación, hectáreas reforestadas, número de comités operando). El procedimiento, estandarizado a través del SGPOA, facilita la síntesis de los resultados obtenidos por la dependencia.

En este tipo de evaluaciones, el significado del concepto “resultados” se relaciona con la obtención de productos tangibles que se asume son predictores del valor público generado. Sin embargo, para confiar en los indicadores que miden la ejecución de actividades como una métrica de impacto, sería necesario que existiera una alta predictibilidad entre la acción y el efecto esperado. Esto no sucede así con el manejo de una ANP, donde la causalidad no es previsible de forma lineal. Por ello, la evaluación del POA no genera información sobre el cambio producido. Para cubrir esta limitación en la herramienta, parece un paso en la dirección correcta la propuesta de incorporar indicadores de efectividad en el SGPOA.

Se ha probado que, más allá de la metodología, se necesita la confluencia de recursos, tiempo y capacidad de decisión para hacer y usar evaluaciones (Cousins *et al.*, 2008; Kusters *et al.*, 2011). Por ello, es crucial complementar la incorporación de indicadores con sistemas organizacionales que soporten su medición. Esto es relevante porque, dada la brecha presupuestal de la dependencia y las restricciones impuestas por procedimientos administrativos, los guardaparques del ANP no cuentan con medios para dar seguimiento a las acciones. A pesar de ello, la evaluación y reelaboración del POA juega un papel importante en su proceso de aprendizaje. Los guardaparques acumulan una gran cantidad de experiencias y las utilizan para tomar decisiones constantemente. De esa manera, aunque no ofrezca evidencia rigurosa, el procedimiento fomenta espacios de intercambio que permiten al grupo entender qué es lo que pasa con la implementación de los proyectos. El proceso es informal e imperfecto, pero útil para orientar decisiones colaborativas.

Para que este proceso de planeación estratégica ocurra, tiene un papel fundamental el liderazgo del director de ANP. En buena medida, de él depende la orientación hacia objetivos de largo plazo en el manejo del área. Investigaciones sobre el tema han concluido, en un sentido similar, que cuando no es posible evaluar el impacto del trabajo realizado, es más relevante valorar el liderazgo y la disposición del personal por compartir las dificultades que enfrentan, así como su capacidad de aprendizaje (Guijt, 2008; Raeside, 2011).

Por este motivo, las metodologías de evaluación de la efectividad que utiliza la dependencia son pertinentes. El propósito de metodologías como INDIMAPA o i-efectividad se orienta hacia el aprendizaje, al identificar aspectos susceptibles de mejora en el manejo de las ANP. Como resultado, los instrumentos tienen la intención de incorporar prácticas gerenciales que incrementen la efectividad del equipo de guardaparques. Sin embargo, sería un error usar este tipo de evaluación para demostrar el impacto de las acciones, pues los puntajes obtenidos no son predictores confiables de la obtención de resultados de conservación (Carranza *et al.*, 2014). A su vez, asumir que la información de estas evaluaciones será utilizada para mejorar el desempeño puede ser una idea equivocada. Como indica Guijt (2008), al hacer evaluaciones se suele hacer énfasis en la información, suponiendo que con ella todo caerá en su lugar, pero no se entienden las relaciones sociales que inciden sobre su uso.

En el caso de estudio, el director del ANP considera que la información del INDIMAPA ha sido útil para que la DES haga un diagnóstico global de la situación de las ANP en el país, y no para tomar decisiones directamente en el área. La implementación de la evaluación i-efectividad se encontraba en marcha cuando se hizo la investigación, por lo que no se cuenta con datos sobre su influencia. Es importante que el uso de este tipo de evaluaciones sirva para hacer cambios en las áreas, y no sólo para que la dependencia compruebe el cumplimiento de la meta 11 de Aichi. Para ello, los guardaparques necesitan concebir la herramienta como un medio para apoyar sus esfuerzos, y no como un procedimiento para reportar información hacia arriba que se suma a su carga de trabajo.

En este sentido, por sus efectos sobre el comportamiento, se ha destacado la influencia que tiene la percepción de las consecuencias que podrían derivarse de las evaluaciones (Ebrahim, 2005; Guijt, 2008). En específico, se ha documentado la regularidad con la que las evaluaciones de desempeño son distorsionadas cuando usan métricas cuantitativas. Esto es así, pues causan que aquellos que son valorados ajusten sus decisiones al cumplimiento de los indicadores (Campbell, 1976; Bovens *et al.*, 2008). En la dirección de ANP, el temor a las consecuencias negativas de la evaluación del POA ha generado el desplazamiento de objetivos. De esa manera, en ocasiones los guardaparques han procurado el cumplimiento de metas anuales, aunque eso no implique necesariamente la obtención de resultados de largo plazo. Es posible que esta presión se replique a nivel directivo si se tiene la expectativa de que la dependencia cumpla con metas anuales predeterminadas.

Lejos de contribuir al manejo adaptativo, estos esquemas erosionan la posibilidad de construir diálogos reflexivos entre los grupos involucrados, al enfatizar la satisfacción de los requerimientos por sobre la necesidad de un mejor desempeño en el mundo real (Cousins *et al.*, 2015). Esta ha sido una crítica recurrente al enfoque de Gestión basada en resultados, especialmente cuando se parte del supuesto de que el cambio se puede proyectar y que toda desviación es responsabilidad del implementador (Eyben, 2008). Dado lo anterior, se ha dicho que en contextos complejos la rendición de cuentas debería hacerse al demostrar el grado en el que se tomaron decisiones basadas en lo aprendido, y cómo esto permitió responder mejor al contexto de la intervención (Perrin, 2002; Guijt, 2010). Cuando ese es el propósito, el valor de una evaluación se mide en función de su capacidad para fomentar procesos colectivos de aprendizaje (Roetman, 2011).

Después de todo, una de las condiciones para lograr un manejo adaptativo es que los guardaparques tengan autonomía para ajustar la operación con base en su conocimiento de

las condiciones del área. Por lo dicho, el uso potencial de la evaluación i-efectividad como herramienta de aprendizaje estaría en función del liderazgo del director de ANP, de su percepción de las consecuencias que traería, y de la capacidad del equipo para llevar a cabo ajustes en su manejo del área.

Conclusiones

Por su cercanía a los proyectos de conservación, los guardaparques son el usuario principal de evaluaciones que fomentan un manejo adaptativo. Esto hace que su punto de vista sea crucial para identificar las fortalezas y debilidades de las evaluaciones empleadas. La investigación analizó la influencia del sistema de evaluación de la CONANP sobre las decisiones de manejo de guardaparques que administran un ANP en el centro de México. Se identificó que éstos interactúan con tres tipos de evaluación: 1) evaluación del POA, 2) evaluación de la efectividad del programa de manejo, 3) evaluaciones de la efectividad del manejo.

La evaluación del POA no involucra la medición del efecto de las acciones. Por ello, de manera formal el procedimiento no genera información para que los guardaparques tomen decisiones con base en el conocimiento de los cambios generados. De cualquier modo, existe una baja capacidad en la dependencia para dar seguimiento a las acciones. En este contexto, el procedimiento tiene una consecuencia positiva para el proceso de aprendizaje de los guardaparques. Al vincularse con la planeación operativa, se convierte en un espacio de reflexión donde fluye información que parte de la experiencia de implementación de los guardaparques. Su posición epistémica preferencial les permite identificar errores y hacer propuestas de mejora.

Dado que estas reuniones de planeación estratégica dependen del liderazgo del director de ANP y lo único que se reporta es el POA finalizado, es recomendable que el procedimiento incorpore la sistematización del aprendizaje obtenido. En específico, los guardaparques podrían asociar teorías del cambio a las actividades de su POA y a los proyectos financiados por medio de programas de subsidio. Este tipo de modelos son un paso esencial para conducir evaluaciones, pues permiten identificar si la lógica causal es correcta. Al contrastar la teoría con la práctica, el modelo facilita la identificación de relaciones y supuestos que no se habían considerado, lo que mejora la comprensión del problema (Margoluis *et al.*, 2013).

La inclusión formal de teorías del cambio como parte de la planeación anual permitiría incorporar indicadores de impacto y asegurar que el aprendizaje es retenido por la organización. Sin embargo, el valor del procedimiento se perdería si no existe apertura organizacional hacia el reporte de fallas, o cuando los guardaparques no tienen autonomía de decisión. La evaluación de la efectividad del PM se concentró en la medición del cumplimiento de las acciones comprometidas en el programa. Sin embargo, esta información no permite deducir si se cumplieron sus objetivos de conservación y desarrollo. Por esta razón, es limitado el aprendizaje que pueden derivar los guardaparques. Aun así, la experiencia adquirida al elaborar la cédula de evaluación permitió a los guardaparques identificar áreas de mejora en sus procedimientos para almacenar información y sirvió como insumo para la actualización del PM.

En teoría, este tipo de evaluación se debe realizar cada cinco años. Por ello, sería conveniente incorporar indicadores de impacto en la actualización del programa, como la dinámica de la cubierta vegetal, la variación poblacional de especies en riesgo o la medición de indicadores de bienestar en las localidades. Debido a la extensión del ANP, en ella confluyen múltiples actores que intervienen el territorio. Esto hace que no sea atribución exclusiva de la CONANP la promoción de objetivos de desarrollo sostenible. En este contexto, y dadas las limitaciones presupuestales de la dependencia, sería recomendable conformar una alianza para evaluar el impacto de las acciones.

Las evaluaciones de la efectividad del manejo (INDIMAPA, i-efectividad) reflejan el interés de la dependencia por las evaluaciones como una herramienta de aprendizaje. Sin embargo, por la experiencia con el INDIMAPA, es necesario garantizar un balance entre la generación de información para audiencias externas y el propósito de apoyar, con información, las decisiones de los guardaparques. La herramienta i-efectividad aún se encuentra en proceso de institucionalización, pues se está probando su metodología en áreas piloto. Con el procedimiento aún por consolidarse, se podría esperar que la dependencia se concentre más en hacer la evaluación que en su capacidad para utilizarla. Por ello, sería conveniente hacer un análisis futuro de la influencia de este tipo de evaluación sobre las decisiones de manejo de los guardaparques en ésta y otras ANP.

En el ANP estudiada, algunas barreras a superar se relacionan con el margen de maniobra de los guardaparques para implementar cambios. Por un lado, debido a limitaciones presupuestales que inciden en su capacidad operativa y, por el otro, por las consecuencias negativas de un enfoque evaluativo que exige demostrar resultados cuantificables en el corto plazo.

Capítulo cinco

El uso de evaluaciones en un centro público de investigación agrícola

The use of evaluations in a public agricultural research centre

Eduardo Luna-Sánchez, Mayra Juliana Chávez-Alcalá, Margaret Skutsch

Revista: **Revista de la Educación Superior**; En proceso de evaluación de pares

El uso de evaluaciones en un centro público de investigación agrícola

Eduardo Luna Sánchez,
Mayra Juliana Chávez Alcalá
Margaret Skutsch

Resumen

La investigación estudia el uso de evaluaciones en un Centro Público de Investigación Agrícola. El objetivo central del centro es transferir tecnología para mejorar el manejo de recursos naturales en una reserva de biosfera del centro de México. Para acceder a la información se combinó el análisis de documentos institucionales con un enfoque etnográfico orientado al aprendizaje. Se identificó que las evaluaciones tienen tres propósitos: verificar la calidad de los proyectos antes de su aprobación, rendir cuentas sobre los resultados inmediatos obtenidos y tomar decisiones laborales sobre el personal académico. No se hicieron evaluaciones para estimar los cambios generados, ni para retroalimentar los proyectos con la finalidad de mejorarlos. Para medir el valor público generado mediante los proyectos de investigación agrícola, es necesario reformar el sistema de evaluación utilizado por la universidad a la que pertenece el centro de investigación. En este sentido, se presenta una propuesta de mejora procedimental para la evaluación de proyectos de investigación aplicada.

Palabras clave: Evaluación de impacto, Universidades Públicas, Rendición de cuentas, Interacciones productivas

Abstract

The research studies the use of evaluations in a Public Agricultural Research Center. The center transfers technology to improve the management of natural resources in a Biosphere Reserve of central Mexico. An analysis of institutional documents was combined with an ethnographic approach to gain access to the information. It was identified that evaluations of the work of this institution had three purposes: to verify the quality of projects before their approval, to give an account of the outputs obtained, and to evaluate the performance of researchers. No evaluations were made to estimate the changes generated, or to provide feedback on the projects in order to improve them. The university to which the center belongs needs to reform its evaluation system, so that it measures the public value generated through agricultural research projects. In view of this, a proposal is presented for the evaluation of applied research projects carried out by the university.

Keywords: Impact evaluation, Public Universities, Accountability, Productive Interactions

Introducción

En México, es notoria la brecha entre una minoría de productores que compiten en el mercado agrícola¹ y una mayoría de pequeños productores que recurren a la pluriactividad porque sus ingresos son insuficientes (Grammont, 2010). Además, dado que estos últimos poseen superficies pequeñas y marginales, son especialmente vulnerables a los efectos del cambio climático (Cohn *et al.*, 2017; Harvey *et al.*, 2018). Por ello, se considera esencial desarrollar tecnologías que fortalezcan la sostenibilidad y resiliencia de sus sistemas productivos (InterAcademy Partnership, 2018).

En ese contexto, los Centros Públicos de Investigación Agrícola (**desde ahora, CPIA**) juegan un papel central, pues generan conocimientos que atienden los problemas sociales y ambientales de las áreas rurales (Rivera *et al.*, 2016). En específico, se espera que experimenten, validen y transfieran tecnologías para desarrollar las capacidades de los productores agrícolas (Ekboir, 2009; Ekboir, 2016).

Esos son objetivos complejos, sobre todo en escenarios donde existen menguantes expectativas de vivir de la agricultura (Giger, 2000; Kessler, 2006). Aunque agricultores innovadores adopten las tecnologías, no se puede deducir que se insertarán con éxito en las cadenas productivas (Ekboir, 2009). Por ello, entender los procesos de cambio tecnológico e identificar los resultados obtenidos son conocimientos indispensables para las investigaciones agrícolas orientadas al desarrollo (Glover *et al.*, 2019).

En ese contexto, las evaluaciones de la investigación agrícola pueden utilizarse con dos audiencias en mente. Por un lado, aportan información a los investigadores sobre cómo sus proyectos influyen tanto en el bienestar de los productores como en el manejo sostenible de los recursos. Con este conocimiento pueden mejorar la efectividad de sus proyectos (Ekboir, 2016; Benoit *et al.*, 2016; Weißhuhn *et al.*, 2017). Por otro lado, brindan información a tomadores de decisión sobre los beneficios de los proyectos de investigación financiados (Hall *et al.*, 2003).

En la rendición de cuentas de actividades científicas ya no se considera suficiente que la evaluación se restrinja a la valoración entre pares. Ahora, se espera que las organizaciones académicas se responsabilicen de las consecuencias prácticas del conocimiento que generan. Por ello, se demanda que adopten enfoques orientados a la obtención de resultados (Serrat, 2010; Nowotny *et al.*, 2003; Sarewitz, 2016; Ekboir *et al.*, 2017). En el caso de la investigación agrícola, la evidente falta de adopción del conocimiento ha promovido cambios en los esquemas de financiamiento, pasando del financiamiento a ciegas a la solicitud del cumplimiento de efectos mensurables y previamente acordados (Lepori *et al.*, 2007 en Rivera *et al.*, 2016; Polastri y Savastano, 2020).

Si bien, se confía en que las evaluaciones sirvan para que las organizaciones aprendan², con frecuencia esta expectativa no se cumple. Una de las razones que explican dicha situación es

¹ El concepto “agricultura” se usa de manera amplia, por lo que incorpora la producción agrícola, pecuaria y silvícola.

² El aprendizaje organizacional sucede cuando sus integrantes adquieren la capacidad de detectar, corregir y compartir errores en los supuestos que fundamentan la operación. Como resultado, el personal es capaz de responder de mejor manera al contexto que enfrenta la organización (Argyris y Schön, 1978).

el uso de las evaluaciones con fines exclusivamente sumativos, asumiendo desde una lógica simplista que, si los proyectos son bien ejecutados, deben conseguir los resultados esperados (Mackay y Horton, 2003; Ekboir, 2016).

En México se ha estudiado poco la forma en que se evalúan los proyectos de vinculación e investigación aplicada de universidades públicas³. El énfasis se ha puesto en la influencia de la política científica federal sobre el comportamiento de los investigadores (Galaz y Gil, 2013; Suárez y Muñoz, 2004). Con la finalidad de disminuir esa brecha de conocimiento, el artículo analiza las evaluaciones con las que interactúa un CPIA que opera dentro de una reserva de la biosfera del centro de México. En ese sentido, se incluyen tanto las evaluaciones externas de sus proyectos financiados con recursos públicos, como el sistema interno de evaluación de la universidad a la que pertenece.

Es importante señalar que esta investigación no pretende evaluar el impacto de los proyectos del CPIA. En cambio, busca responder las siguientes preguntas: a) ¿Cómo y para qué se hacen evaluaciones? y b) ¿Qué factores influyen en el propósito de las evaluaciones utilizadas? Debido a que las evaluaciones suelen relacionarse con la reputación de las organizaciones, como un principio ético de la investigación, se tomó la decisión de garantizar el anonimato de la universidad estudiada.

La información se organiza de la siguiente manera. Primero, se detalla la metodología utilizada para responder las preguntas de investigación y después se describe el marco conceptual a partir del cual se estructura el análisis. En la sección de resultados se desarrollan las características de los diferentes tipos de evaluación asociados al CPIA. En la discusión, se contrastan los hallazgos con la literatura sobre el tema, se resaltan las áreas de oportunidad identificadas y se hace una propuesta de mejora procedimental para la evaluación de proyectos de investigación agrícola en la universidad estudiada. Finalmente, en la conclusión se sintetizan los mensajes centrales del artículo.

Metodología

Los estudios de caso son un tipo de investigación cualitativa que analiza fenómenos en su propio contexto, lo que permite relacionar datos empíricos con planteamientos teóricos (Yin, 2009). Por ello, un caso permite contrastar supuestos sobre el valor intrínseco de las evaluaciones con lo que sucede en la práctica.

La metodología seleccionada para abordar el caso es la Perspectiva de Aprendizaje (PdA) propuesta por Eyben (2009). Esta es un enfoque etnográfico para estudiar organizaciones que conducen proyectos de desarrollo. En lugar de limitarse a la descripción de cómo funciona el desarrollo, la PdA asume que existe un área interesante de uso práctico, si se combinan la crítica de la perspectiva etnográfica con la instrumentalidad de aquellos que desean mejorar la práctica de las organizaciones (Eyben, 2009: 84). Por ello, al estudiar una organización, esta metodología se interesa simultáneamente en atender problemas organizacionales,

³ Si bien, toda investigación busca generar un cambio al aportar nuevos conocimientos, se considera que una investigación es aplicada cuando existe una intencionalidad expresa de utilizar los conocimientos generados para atender problemáticas sociales.

promover el desarrollo de capacidades en sus integrantes y generar conocimientos científicos. En este caso, los investigadores interactuaron con el personal de la organización para explorar el uso de evaluaciones y, a la vez, proponer cambios que permitan responder a modelos deseados de desempeño organizacional.

Las técnicas de investigación que se utilizaron fueron:

1. Revisión de documentos. La recopilación de documentos relacionados con el CPIA y las evaluaciones de la universidad se realizó a través de una exploración de su página web institucional. De esa manera, se tuvo acceso a convocatorias, documentos de planeación estratégica y formatos de registro de proyectos de investigación. Además, a través de solicitudes de acceso a la información, se recopiló información sobre proyectos con financiamiento externo.

2. Entrevistas semiestructuradas. Las entrevistas se hicieron a personal directivo y operativo que labora o laboró en la universidad. El criterio de selección de las personas entrevistadas fue la relación con diferentes procedimientos de evaluación. Sin embargo, la selección incluyó personal que se encuentra en diferentes posiciones, dada la posible existencia de diferencias intraorganizacionales. Las entrevistas permitieron recolectar información sobre la puesta en práctica de las evaluaciones, entremezclada con la opinión y reflexiones de los propios entrevistados en torno al tema.

3. Observación participante. A través de ella, fue posible explorar la cotidianidad e interacciones del personal de la universidad. Esto permitió hacer descripciones detalladas de diferentes aspectos relacionados con su cultura organizacional. De manera específica, se asistió a múltiples reuniones trabajo, algunas de ellas facilitadas por los investigadores, y se propusieron procedimientos para reorientar partes del sistema evaluativo del CPIA. De manera particular, se aprovechó la invitación para colaborar en un proyecto para ofertar una evaluación de diseño que mejorara la descripción de sus impactos.

Evaluaciones en centros de investigación

En función del propósito que se tenga, existen diferentes formas de evaluar la investigación agrícola. Por ello, para enmarcar el estudio se presenta una síntesis proveniente de la literatura sobre el tema. Primero, se desarrollan las características de evaluaciones de desempeño, distinguiendo entre enfoques basados en la obtención de productos y enfoques que miden interacciones productivas (Spaapen y Van Drooge, 2011). Acto seguido, se contrastan los objetivos de evaluaciones de impacto y evaluaciones formativas, señalando sus fortalezas y limitaciones (Cuadro 1).

La selección de uno u otro tipo de evaluación depende de factores ligados al solicitante, el cual define cómo, por quién y para quién se conduce una evaluación. Por ello, el contexto organizacional y los arreglos institucionales juegan un papel clave (Mackay y Horton, 2003; Hall *et al.*, 2003; Taut, 2007).

Cuadro 1. Diferentes tipos de evaluación asociados a la investigación agrícola.

	Evaluaciones de desempeño	Evaluaciones de impacto	Evaluaciones formativas
Propósito	Rendición de cuentas Aprendizaje organizacional	Rendición de cuentas	Aprendizaje organizacional
Unidad de análisis	Outputs ⁴ <ul style="list-style-type: none"> • Productos académicos • Productos de intermediación (talleres, conferencias, material de divulgación) • Resultados inmediatos de actividades (infraestructura, área reforestada, etc.) Outcomes ⁵ <ul style="list-style-type: none"> • Efectos de la intermediación de conocimientos 	Outcomes <ul style="list-style-type: none"> • Rendimientos • Retorno de la inversión • Tasa de adopción 	Outcomes <ul style="list-style-type: none"> • Procesos • Procedimientos organizacionales

Fuente: Elaboración propia

En México, una parte de la investigación agrícola se realiza en Universidades Públicas Estatales. Estas son organismos públicos descentralizados que cuentan con autonomía institucional para llevar a cabo las funciones de docencia, generación y aplicación del conocimiento. Por ende, son capaces de decidir por sí mismas las características de su sistema de evaluación científica (Olvera *et al.*, 2009).

Sin embargo, cuando reciben recursos externos para la ejecución de proyectos, precisan ceñirse a los procedimientos evaluativos de donantes, normalmente entidades gubernamentales. Un porcentaje importante de este financiamiento proviene del Consejo Nacional de Ciencia y Tecnología⁶ (CONACYT), dependencia que maneja fondos competitivos para promover proyectos (Vera *et al.*, 2008; Rivera *et al.*, 2016).

Además, CONACYT otorga recursos directamente a los investigadores, ya sea a través de programas de pago por mérito que incentivan su productividad académica (Galaz *et al.*, 2008; Gil, 2014) o a través de programas que financian su trabajo en Cuerpos Académicos⁷, los cuales implementan proyectos con la intención de “transformar su entorno” (SEP, 2018, pág. 16).

En este contexto, especialmente si existe una dependencia hacia recursos externos, cabe la posibilidad de que estas reglas e incentivos, establecidos a nivel federal, influyan tanto en los

⁴ Son los productos concretos y contables que indican la realización exitosa de las acciones de un proyecto.

⁵ Se refieren a los cambios que se espera sean ocasionados gracias a la obtención de los productos de una intervención.

⁶ Organismo público responsable de articular las políticas públicas ligadas a la investigación científica en México.

⁷ Los Cuerpos Académicos son grupos de profesores de tiempo completo que comparten una o varias líneas de generación y aplicación del conocimiento (SEP, 2018).

mecanismos de evaluación empleados por las universidades como en las prioridades de los académicos.

Evaluaciones de desempeño

Para evaluar los resultados inmediatos de proyectos de investigación, con frecuencia se recurre a productos académicos como publicaciones en revistas de prestigio y presentaciones en congresos. Este enfoque, predominante en la política científica latinoamericana, asume que los artículos sujetos a un proceso de arbitraje y los índices globales de citas miden adecuadamente la calidad e influencia de las investigaciones (Vasen y Lujano, 2017). Después de todo, se espera que los investigadores generen conocimientos, por lo que parece lógico que su efectividad se mida con los medios a través de los que difunden sus hallazgos.

No obstante, se ha señalado que ese tipo de indicadores tienen múltiples problemas cuando se utilizan como única medida de evaluación (Ekboir, 2016; Vasen y Lujano, 2017; Chavarro y Ràfols, 2018). En campos naturalmente inclinados hacia la investigación aplicada, como las ciencias agrícolas o ambientales, es necesario que el conocimiento incida en el comportamiento de otros actores (Netherlands Organization for Scientific Research, s.f.). Sin embargo, no se puede dar por hecho que los grupos involucrados accederán, entenderán y utilizarán los artículos científicos para tomar decisiones (Salafsky *et al.*, 2001). En el fondo, si solo se utilizan indicadores ligados a la productividad académica, la valoración de los investigadores se fundamenta en cifras y no en juicios que analizan el efecto del conocimiento generado (Hicks *et al.*, 2015).

En las ciencias agrícolas, se ha señalado que la mayor parte de las publicaciones no proveen soluciones para las necesidades de pequeños productores locales (Nature, 2020). Esto ocurre, en parte, por las prioridades del financiamiento internacional en investigación agrícola, y por tendencias en la agenda de investigación internacional (Ekboir *et al.*, 2009; Nature, 2020). Los investigadores, interesados en publicar en revistas de prestigio para avanzar en su carrera académica, encuentran que el sistema de evaluación menosprecia actividades de intercambio fuera del ámbito científico, lo cual entorpece la colaboración con otros actores clave para la innovación agrícola (Ekboir, 2016; Vera y Dutrénit, 2016; Rivera *et al.*, 2016).

Para superar las limitaciones anteriores, se han desarrollado evaluaciones que valoran las interacciones entre investigadores y usuarios del conocimiento (Spaapen y Van Drooge, 2011). Estos enfoques analizan los mecanismos de intercambio utilizados por el investigador y buscan pruebas de que su aporte al conocimiento influyó en la toma de decisiones de otro actor u organización (LSE Public Policy Group, 2011). En ese sentido, los indicadores monitoreados van más allá de los medios a través de los cuales se mantuvo contacto con los grupos interesados. Asimismo, se incluye evidencia del uso del conocimiento o el grado en que los grupos interesados valoran los resultados de la investigación (Spaapen *et al.*, 2011).

El supuesto del que parten tales enfoques evaluativos es que la intermediación y el uso de conocimientos son condiciones necesarias para que el impacto pueda ocurrir, por lo que pueden utilizarse como indicadores *proxy* del impacto de una investigación (Spaapen y van Drooge, 2011). Además, se considera que el análisis del proceso arroja información valiosa para retroalimentar las investigaciones, de forma que sirven simultáneamente para rendir cuentas y aprender (Ekboir, 2003). En ese sentido, pueden incorporarse reportes de autoevaluación donde los investigadores reflexionen sobre el cambio conseguido y especifiquen ajustes para obtener mejores resultados (Van der Meulen *et al.*, 2010).

Este enfoque evaluativo implica ajustar el rol esperado del científico, de forma que no solo se comunique con sus pares, sino con otros actores del sistema de innovación agrícola (Ekboir, 2016). Por ello, además de generar evidencia y probar nuevas tecnologías, los equipos de investigación precisan habilidades para facilitar actividades de intermediación del conocimiento. Por ejemplo, saber facilitar procesos, organizar reuniones efectivas y trasladar la demanda de conocimiento de los actores a preguntas científicas (Turnhout *et al.*, 2013).

Evaluaciones de impacto

A nivel internacional, la baja efectividad de los proyectos de investigación agrícola promovió la adopción de evaluaciones que demostraran el valor público generado. El paradigma dominante desde la década de los 80 fue la estimación de los impactos económicos a través de cálculos de costo-beneficio (Ekboir, 2003, 2016; Benoit *et al.*, 2016; Weißhuhn *et al.*, 2017).

Existen distintas metodologías para estimar los beneficios generados. Por un lado, se encuentran estudios de adopción tecnológica que identifican el grado en el que las innovaciones son utilizadas. Esto permite medir su efecto sobre la productividad o los ingresos. Por el otro, se encuentran métodos econométricos que aíslan el efecto del conocimiento científico sobre la productividad, con la finalidad de calcular el retorno de la inversión en proyectos de investigación (Hall *et al.*, 2003).

Se han identificado algunas limitaciones de este enfoque evaluativo. En primer lugar, si bien se puede calcular el impacto económico de investigaciones relacionadas con la innovación y la transferencia de tecnología, es muy difícil estimarlo en el corto plazo por problemas de atribución o de latencia (OSIRIS, 2016). Esta situación se acentúa si consideramos que la norma es evaluar proyectos que han tenido entre uno y tres años de duración, los cuales no vuelven a evaluarse tiempo después.

En segundo lugar, se considera errónea su visión simplificada del proceso de cambio, al asumir una relación directa entre financiamiento (*input*), tecnología innovadora (*output*), adopción (*outcome*) e incrementos en las ganancias (*impacto*). El problema es que, al enfocarse solo en el “punto de llegada”, se soslaya el contexto de las investigaciones (Ekboir, 2003). Normalmente, estas tienen un alto grado de incertidumbre, por ende, si se solicitan resultados en plazos determinados se puede desincentivar el surgimiento de proyectos novedosos que asuman mayores riesgos, a veces fallen, pero que tengan el potencial de generar beneficios en el largo plazo (Ekboir, 2016).

Finalmente, se considera que el sesgo de estos enfoques evaluativos hacia resultados económicos coloca dentro de cajas negras los procesos y consecuencias del cambio tecnológico. En este sentido, se ha criticado el uso simplista de la tasa de adopción como indicador de efectividad. Este concepto, normalmente considerado en términos binarios⁸, anula la agencia del usuario de la tecnología, al asumir que un cálculo racional debería conducir a su selección. Sin embargo, los usuarios utilizan múltiples criterios para adaptar la tecnología a sus aspiraciones, modificando aspectos de esta con base en un proceso dinámico de aprendizaje y experimentación (Glover *et al.*, 2019).

⁸ Es decir, usar o no determinados materiales, herramientas y métodos.

Además, medir la adopción no registra las ramificaciones del cambio tecnológico, incluyendo impactos no intencionados, tanto positivos como negativos (Glover *et al.*, 2019). Por lo anterior, este tipo de evaluaciones tienen una relación escasa con objetivos de aprendizaje organizacional que se basan en la comprensión de procesos (Hall *et al.*, 2003; Weißhuhn *et al.*, 2017).

Evaluaciones formativas

Este tipo de evaluación, basada en casos, complementa la identificación de los efectos de las investigaciones con un análisis de los procesos involucrados en su generación (Mackay y Horton, 2003; Benoit *et al.*, 2016). Su propósito es retroalimentar la forma de trabajar para incrementar la probabilidad de obtener los resultados deseados (Ekboir *et al.*, 2017).

El enfoque hace énfasis en las interacciones, agencia y capacidades de múltiples actores involucrados en el desarrollo y uso de las tecnologías (Glover *et al.*, 2019). Además, se basa en la combinación de cadenas de resultados y actividades de monitoreo, con la finalidad de identificar desviaciones e implementar acciones correctivas (Guijt, 2008; Alvarez *et al.*, 2010; Ekboir, 2016).

Una característica importante de este tipo de evaluaciones es la promoción de capacidades de aprendizaje en los equipos de trabajo, lo cual implica crear oportunidades para que los investigadores reflexionen, intercambien conocimiento y transformen su práctica (Byrne y Vincent, 2004). En este sentido, los proyectos “fallidos” no se ocultan, sino se valoran como una fuente de información (Ekboir, 2016). Esto implica revisiones introspectivas de la forma en la que la comunidad académica opera, así como de las interacciones que tiene con su contexto institucional (Hall *et al.*, 2003). Por estos motivos, tales evaluaciones se han propuesto como parte esencial del modelo de gestión de centros de investigación (Ekboir, 2016).

Descripción del caso

Diferentes rankings posicionan a la universidad pública estudiada como una de las mejores del país, en términos de su productividad académica. En diferentes documentos institucionales de las dos últimas administraciones (2012-2021) se hace referencia a la investigación como una actividad que refleja el compromiso de las universidades públicas por resolver problemas de su entorno. Por tanto, tiene sentido que su modelo de gestión organizacional sea el de la Responsabilidad Social Universitaria, enfoque que busca manejar los impactos generados por la universidad (Vallaey, 2014).

Una de las manifestaciones visibles de este interés ha sido el reconocimiento de actividades de vinculación realizadas por sus académicos. Por ello, la universidad cuenta con procedimientos que financian y recompensan proyectos que busquen mejorar las condiciones del entorno. Otra política universitaria que refleja su interés en procesos de desarrollo ha sido el establecimiento de campi en varios municipios del estado. La intención ha sido acercar licenciaturas *ad hoc* y promover investigaciones que atiendan problemas locales.

El CPIA analizado es un ejemplo de esta vocación, pues se estableció en una zona históricamente marginada del territorio estatal, la cual, fue declarada como reserva de la biosfera por su patrimonio natural y cultural. Para construir el CPIA, en 2013, la universidad concursó en un programa de financiamiento que unifica recursos del CONACYT y el gobierno estatal. El fideicomiso, conocido como Fondo Mixto, tiene dos propósitos: apoyar la investigación científica orientada a la solución de problemas locales y consolidar las capacidades científicas de los estados (Rivera *et al.*, 2016; Rocha y De la Cerda, 2016).

Los términos de referencia de la convocatoria solicitaban la creación de un centro de investigación que, a través de la transferencia tecnológica, mejorara los procesos productivos de la región del estado en donde se encuentra la reserva de la biosfera⁹. En 2014 el proyecto recibió el financiamiento y el CPIA comenzó a construirse en un terreno de 21 ha. donado por uno de los municipios de la región.

Actualmente, el CPIA cuenta con cuatro módulos especializados en el ámbito pecuario, agrícola, acuícola y forestal. El modelo de gestión del CPIA se basa en un enfoque multidisciplinario, colaborando en éste diez profesores-investigadores con perfiles complementarios (ciencias biológicas, agropecuarias y sociales) (Cuadro 2).

Cuadro 2. Módulos del CPIA.

Módulo	Proyectos	Perfil del coordinador
Pecuario	-Asesoría en ovinocultura. -Alternativas de nutrición de rumiantes. -Desarrollo de sistemas silvopastoriles.	Veterinario con Maestría en Producción Animal Sustentable
Agrícola	-Plantaciones experimentales de <i>Jatropha</i> para la producción de biodiesel. -Capacitación en agricultura protegida. -Área demostrativa y experimental de producción de hortalizas.	Ingeniero Agrónomo con maestría en Ingeniería de biosistemas
Acuícola	-Capacitación en producción piscícola. -Investigación para reproducir en cautiverio especies nativas con potencial productivo acuícola.	Biotecnólogo con maestría en ingeniería de biosistemas y doctorado en Ingeniería
Forestal	-Reproducción de plantas nativas para rehabilitar áreas forestales degradadas. -Producción de bioinsumos. -Manejo integrado de plagas forestales y frutícolas.	Biólogo con maestría en Recursos Bióticos

Fuente: Elaboración propia.

⁹ En la Reserva se han documentado problemáticas vinculadas con la degradación ambiental y con la falta de un relevo generacional productivo. La migración, estrategia histórica para conseguir la manutención familiar, ha generado un proceso de desagrarización de la economía local, reduciendo las expectativas de vivir de actividades agropecuarias, en el mediano y largo plazo.

Resultados

En esta sección se presentan las distintas evaluaciones asociadas a la operación del CPIA. En primer lugar, se describen las evaluaciones externas que formaron parte de la gestión del proyecto financiado por el Fondo Mixto CONACYT-Gobierno Estatal. Después, por congruencia lógica, se presenta la propuesta de evaluación interna elaborada por la propia universidad. En ambos casos, el objeto de la evaluación fue el proyecto global para establecer el CPIA.

De ahí, se transita hacia los procedimientos que emplea la universidad para evaluar los proyectos específicos que generan los investigadores que forman parte del CPIA. Finalmente, por su influencia sobre el comportamiento de los investigadores, se incluye también una descripción del sistema de evaluación con el que la universidad valora el desempeño de su personal académico. La diferencia entre ambos es el objeto de evaluación. En el primero de los casos, el énfasis está en el proyecto registrado; en el segundo, está en el personal de la organización. El cuadro 3 sintetiza la información que se presentará.

Cuadro 3. Evaluaciones asociadas a la operación del CPIA.

EVALUACION	OBJETO	TIPO DE EVALUACION	PROPOSITO
Fondo Mixto	Proyecto del CPIA	Evaluación <i>ex ante</i> de diseño	Verificación de estándares de calidad
		Evaluación <i>ex post</i> de desempeño	Rendición de cuentas con base en resultados inmediatos
Plan de desarrollo del CPIA	Proyecto del CPIA	Evaluación formativa	Aprendizaje organizacional
Investigaciones	Proyectos de investigación registrados	Evaluación <i>ex ante</i> de diseño	Verificación de estándares de calidad
		Evaluación <i>ex post</i> de desempeño	Rendición de cuentas con base en resultados inmediatos
Ingreso y Permanencia de Personal Académico	Investigadores	Evaluación de desempeño	Toma de decisiones de contratación y promoción laboral

Fuente: Elaboración propia.

Fondo Mixto

En las reglas de operación de Fondo Mixto se establecen los criterios de evaluación con los que se determinará la calidad técnica de las propuestas (*ex ante*) y los resultados generados (*ex post*). Con respecto de la primera, las reglas de operación establecen dos valoraciones: la primera, en términos de pertinencia, verifica que los productos entregables comprometidos atiendan las expectativas de la demanda establecida en el Fondo Mixto; la segunda evalúa la calidad científica de la propuesta. Para ello, se valora su carácter innovador, la congruencia entre actividades, productos, metas y las capacidades del grupo de trabajo.

Una vez financiado el proyecto del CPIA, se realizaron evaluaciones de desempeño intermedias, con el objetivo de verificar los avances, y una evaluación final que revisó la entrega de una lista de productos comprometidos (Cuadro 4). Como se ha dicho, la demanda del fondo mixto solicitó la construcción de un centro público de investigación y desarrollo tecnológico que diera respuesta a los requerimientos del sector agropecuario de la región.

Cuadro 4. Algunos productos entregables comprometidos en el proyecto del CPIA.

- Un plan estratégico y organizacional del centro de investigación.
- Un calendario de capacitación continua y un manual de buenas prácticas del área agrícola, pecuaria, acuícola y forestal.
- Un grupo multidisciplinario para asesorar a pequeños y medianos productores de la región.
- Una estrategia de operación para atender adecuadamente las demandas de los productores.
- Un manual de organización y de procedimientos.
- Un programa de formación de recursos humanos
- Publicación de 8 artículos científicos arbitrados.
- Formación de recursos humanos (tesis).
- Unidades demostrativas en producción pecuaria, forestal, acuícola y agricultura protegida
- Cuatro laboratorios especializados.

Si bien, la convocatoria del Fondo Mixto solicitaba, entre los entregables, la formalización de vinculaciones con empresas, organizaciones gubernamentales y otros centros de investigación, sin embargo, en el convenio final no quedaron plasmados estos productos. En cambio, en la propuesta elaborada por la universidad, se incluyeron productos académicos como parte de los entregables (tesis, publicaciones).

Con respecto de la evaluación de impacto de los proyectos financiados por el Fondo Mixto, las reglas de operación establecen que su propósito es conocer la efectividad del programa al medir el beneficio generado con las inversiones. El manual de procedimientos del Fondo establece que la valoración de los beneficios puede ser tanto prospectiva como retrospectiva, preferentemente, después de tres años del término del proyecto. Sin embargo, este tipo de evaluación no es obligatoria, siendo decisión del secretario técnico del fideicomiso si se asigna un presupuesto para ello.

Aun así, los términos de referencia de cada convocatoria del fondo establecen los indicadores de impacto con los que se realizaría la evaluación. De esta manera, los términos de referencia de la convocatoria 2013 establecieron una serie de indicadores para la demanda a la cual postuló el proyecto del CPIA (Cuadro 5).

Además de esos indicadores, el proyecto sometido al Fondo Mixto por la universidad contemplaba métricas adicionales para establecer el impacto. En la propuesta, se estimaba que el CPIA incrementaría en el mediano plazo los ingresos de 4,000 ganaderos de doble propósito, 15,000 productores agrícolas, 850 acuicultores y 1000 productores forestales. Esto implicaba conseguir un aumento de 20,000 toneladas en la producción agrícola, 35,000 toneladas de carne en canal y 500 toneladas anuales de producción acuícola, adiciones que sumarían más de 250 millones de pesos. Por otro lado, se estipuló que el CPIA contribuiría en la conservación de 100,000 hectáreas forestales de la región. Independientemente de estas

métricas, a la fecha no se ha realizado ninguna evaluación del impacto que han tenido las actividades del CPIA.

Cuadro 5. Indicadores para establecer el impacto de la operación de un Centro de Investigación y Desarrollo Tecnológico.

- Número de proyectos de investigación científica relacionados con el desarrollo de la región.
- Número de empleos que se generan con la creación del centro.
- Número de redes estatales de ciencia y tecnología en las que participa el centro.
- Número de redes temáticas nacionales de ciencia y tecnología en las que participa el centro.
- Nivel de autosuficiencia financiera.
- Número de programas científicos transmitidos en radio y televisión
- Número de artículos de divulgación en revistas electrónicas
- Número de publicaciones especializadas
- Participación de estudiantes y docentes en actividades científicas (congresos, seminarios).

Fuente: Demandas específicas del Fondo Mixto convocatoria 2013

Plan de desarrollo del CPIA

Entre 2015 y 2018 la Dirección de Planeación de la universidad reestructuró la forma en la que se realizaba la planeación institucional al interior de esta. Se consideró que los Planes Institucionales de Desarrollo, promovidos por programas de subsidio federal, solo generaban información para comunicar al exterior, sin que fueran útiles para retroalimentar el quehacer universitario. Por ello, se promovieron planes de desarrollo participativos en diferentes escalas, esperando que las evaluaciones servirían simultáneamente para verificar la obtención de resultados y para adaptar continuamente los planes.

Siguiendo lo anterior, en 2016 se promovió que los investigadores desarrollaran un plan trianual de desarrollo del CPIA, mismo que permitiera incrementar su impacto en el entorno. En concreto, el plan debía contener los mecanismos de vinculación entre el CPIA y los problemas de su zona de influencia. Por ello, en el plan resultante se incorporaron actividades de los módulos como la elaboración de diagnósticos de sistemas productivos, el desarrollo de paquetes tecnológicos y la implementación de talleres de fortalecimiento de capacidades con pequeños productores.

El plan estableció indicadores para medir la obtención de resultados inmediatos. Además, estipulaba que se realizaría una evaluación anual para verificar el cumplimiento de las acciones, reflexionar sobre su efectividad y rediseñar tanto las estrategias como los indicadores de evaluación. Sin embargo, el plan de desarrollo del CPIA nunca fue evaluado, ni por los propios investigadores, ni por la Dirección de Planeación de la universidad.

Proyectos de investigación registrados

En la universidad, el registro de proyectos de investigación se basa en dictámenes entre pares para evaluar *ex ante* su validez científica. Para ello, se maneja un protocolo estandarizado que solicita indicar las consecuencias prácticas del conocimiento a generar. En este sentido,

pide plantear la aplicabilidad de la investigación, pensando en su relevancia social. Por otro lado, el protocolo es dictaminado por un comité de ética de la investigación, instancia que valora el cumplimiento de estándares que garanticen la protección de los actores involucrados¹⁰.

Por su parte, la evaluación *ex post* de los proyectos registrados se concentra en revisar la entrega de productos académicos comprometidos (artículos, tesis, patentes, ponencias). Ligado a lo anterior, en 2019 comenzó a operar un procedimiento para incentivar la generación de cierto tipo de productos. Al momento de registrar proyectos de investigación, los investigadores tienen una descarga de más horas si comprometen publicaciones en revistas indizadas reconocidas por su impacto (JCR o SCOPUS), si dirigen tesis de doctorado o si registran patentes. En comparación, su descarga es menor si publican en revistas arbitradas, dirigen tesis de licenciatura o realizan ponencias por invitación.

A su vez, la universidad permite a los investigadores el registro de proyectos de vinculación con financiamiento. En estos proyectos, se debe explicitar la búsqueda de beneficios para actores específicos. En cuanto al procedimiento de evaluación, la universidad no realiza distinciones con respecto de los proyectos de investigación, utilizando el mismo protocolo estandarizado y requiriendo la entrega de algún producto académico como resultado.

En relación a la estimación del impacto, la universidad reconoce las limitaciones de su sistema de evaluación. Una revisión de la estrategia para orientar la investigación hacia la resolución de problemas sociales, contenida en el Plan Institucional de Desarrollo 2015-2018, concluyó que la institución no cuenta con mecanismos para medir el impacto de los proyectos.

Desempeño del personal

La universidad utiliza un sistema cuantitativo a través del cual valora el desempeño académico de su personal. Con base en este, se busca dar una base objetiva a las convocatorias a través de las que los investigadores concursan por plazas de tiempo completo, reciben estímulos y obtienen promociones laborales. El sistema también valora aspectos ligados a la docencia, como cursos de formación pedagógica. No obstante, destacan por su valor, criterios ligados a la productividad académica tales como publicaciones, direcciones de tesis, ponencias en congresos y registros de propiedad intelectual.

Uno de los objetivos del sistema es promover en los investigadores un comportamiento alineado a los indicadores con los que se evalúa el desempeño de la universidad en programas de gobierno que le brindan fondos extraordinarios. Un ejemplo es el Programa para el Fortalecimiento de la Calidad Educativa (*PFCE*), el cual utiliza indicadores de capacidad académica para estimar los avances de las universidades en relación al desempeño de sus académicos de tiempo completo. En específico, el programa valora que los investigadores cuenten con posgrado, sean parte de Cuerpos Académicos y formen parte de esquemas de transferencias monetarias condicionadas, como el Programa de Desarrollo del Profesorado o el Sistema Nacional de Investigadores (Vidaurre y Luna, 2017). Como resultado, en la convocatoria de plazas de tiempo completo, un porcentaje importante de la evaluación se

¹⁰ Como la presentación de su consentimiento previo, libre e informado.

vincula a métricas del Programa de Desarrollo del Profesorado¹¹. Esto con la finalidad de garantizar que el investigador podrá participar sin inconvenientes en la convocatoria, una vez que obtenga la plaza.

La mayor parte de los investigadores del CPIA no tienen una plaza de tiempo completo, por lo que tienen presentes las reglas de las convocatorias de ingreso. Algunos de ellos reconocieron la influencia del sistema, tanto sobre el comportamiento individual, como en los equipos de trabajo a los que pertenecen. En la búsqueda de certidumbre laboral y de mejores ingresos, los investigadores identificaron desafíos para balancear las expectativas de tener proyectos de vinculación con productores locales y, simultáneamente, una alta productividad académica.

“Queramos o no, estamos ceñidos a los indicadores de productividad que nos están pidiendo... Tenemos que seguirlos si queremos seguir avanzando en la universidad.”
(Investigador #1, comunicación personal)

En el año 2019 se tuvo una reunión, facilitada por investigadores de otro campi, con el objetivo de conformar un equipo de trabajo que colaborara en torno de un Cuerpo Académico que se pretendía formar. Una de sus intenciones era la participación grupal en proyectos conjuntos de investigación, de forma que puedan generarse colaboraciones que resulten en publicaciones compartidas. La idea fue presentada como un mecanismo para hacer carrera profesional dentro de la universidad, al cumplir con uno de los principales indicadores de desempeño: la productividad académica.

Como parte de la propuesta, se compartieron algunos acuerdos de integridad para el funcionamiento del Cuerpo Académico. En el documento se señala que los integrantes tienen el derecho a la participación en proyectos de investigación y dirección de tesis, que para ser coautor, al menos, se debe haber participado en la redacción de un 20% del texto y que, en torno de los proyectos de investigación, se tiene que incluir a todos los Profesores de Tiempo Completo. Al momento de la investigación, esta propuesta de organización no había sido retomada por los investigadores del CPIA.

Discusión

La universidad, aunque goza de autonomía institucional, ha transferido el enfoque evaluativo de la política científica federal a sus procedimientos de contratación y promoción laboral. Esto coincide con lo señalado para otras universidades del país (Gil, 2014; Gil y Contreras, 2017). Esta decisión es comprensible, pues los fondos extraordinarios cubren parte de la brecha financiera desde la que operan. La universidad destina el 100% del subsidio ordinario (federal y estatal), y buena parte de los recursos que genera por sí misma, para pagar sueldos y prestaciones de su personal. En ese contexto, los programas concursables de la federación son cruciales para financiar su desarrollo.

¹¹ El programa tiene el objetivo de incentivar a los Profesores de Tiempo Completo, de manera que alcancen un “perfil deseable” al alcanzar, entre otros criterios, el nivel de productividad académica requerido.

Si bien, es útil contar con estándares para tomar decisiones objetivas, la producción académica, medida en términos cuantitativos, tiene limitantes. No es posible ligar la cantidad de investigaciones con el impacto que se ha tenido en el entorno (Bensing *et al.*, 2013). Por ello, si la evaluación de desempeño empleada por la universidad es útil para promover comportamientos deseados en los investigadores, no lo es para establecer la efectividad de las investigaciones.

Uno de los mayores riesgos es que se confundan los indicadores (productos académicos) con los objetivos (resolver problemas sociales), incentivando que se publique como una meta en sí misma (Campbell, 1976). Otras investigaciones han mencionado que esto puede provocar un comportamiento administrativo estratégico en los académicos, mismo que los lleve a destinar más tiempo en cumplir con requisitos para tener estabilidad laboral o generar ingresos adicionales, que en la ejecución de proyectos de largo plazo (Suárez y Muñoz, 2004; Galaz y Gil, 2013; Alonso, 2014).

En el caso de la universidad, investigaciones previas han reconocido que existe una desvinculación entre la producción de conocimientos y la mejora de las condiciones de la reserva de la biosfera. En este sentido, algunos investigadores del CPIA han reconocido una tensión entre mantener una alta productividad, objetivo que les sirve en su carrera académica, y dedicar tiempo a actividades orientadas a problemas locales, esfuerzo que no podrán publicar en las revistas especializadas de su disciplina. Un aspecto a destacar de la reunión en donde se promovió la conformación de un Cuerpo Académico en el CPIA, es que en ella se habló sobre la importancia laboral de incrementar el número de publicaciones, pero nunca se mencionó la relevancia social que estas tendrían en la reserva, en caso de conseguirlo.

Otros estudios también han señalado el riesgo de que la vinculación no sea una prioridad de los académicos mexicanos (Galaz *et al.*, 2008). Sin embargo, la universidad cuenta con varias políticas que buscan modificar esta tendencia al revalorizar proyectos que usen el conocimiento para atender problemas prácticos y actuales, a escala local y regional. Por un lado, permite el registro de proyectos de vinculación financiados, los cuales tienen el mismo valor que los proyectos de investigación en los tabuladores de evaluación académica. Por el otro, a través de los campi regionales, promueve la inmersión de los investigadores en el entorno que se desea intervenir. Aun así, la universidad recurre a los mismos indicadores de productividad académica para evaluar los resultados de proyectos de investigación y de vinculación. Como ya se ha dicho, esto es insuficiente para estimar los efectos de proyectos de conservación o de transferencia tecnológica.

Por su parte, la evaluación externa del proyecto para establecer el CPIA, financiado por el Fondo Mixto, basó la rendición de cuentas en la obtención de productos previamente acordados (infraestructura, productos académicos, documentos de planeación y operación). Evidentemente, estos indicadores son insuficientes para establecer el valor público generado. Recientemente, la dirección del CONACYT cuestionó la calidad de las evaluaciones de los Fondos Mixtos. Esto al no poder identificar si servían para resolver los problemas señalados en las demandas establecidas en las convocatorias (Clemente, 2019).

Por problemas de latencia y atribución, es difícil evaluar en el corto plazo el impacto de proyectos relacionados con innovaciones en el sistema agrícola (Ekboir, 2016). Por ello, se considera prioritario garantizar que los proyectos tengan un diseño causal riguroso, el cual permita dar seguimiento a cambios que, se considera, sean condiciones necesarias para tener éxito. De ahí el interés del campo del desarrollo por los modelos lógicos, pues estos

representan la causalidad, permiten identificar supuestos y planificar el monitoreo (Margoluis *et al.*, 2013).

Como parte de estándares mínimos de evaluaciones *ex ante*, estas teorías del cambio servirían tanto al CONACYT como a la universidad para evitar la entrega de recursos a proyectos que exageren, sin evidencia alguna, sus posibles efectos. Esto se debe, en parte, a la responsabilidad de los donantes, ya que es una práctica frecuente que las organizaciones, para asegurar el financiamiento, presenten propuestas románticas que prometen generar profundos cambios sociales, a pesar de que no sea posible conseguirlos con el tiempo y los recursos disponibles (Guijt, 2008). En el caso del CPIA, sin que mediara una cadena de resultados que lo justificara, se estimó un aumento de 35,000 toneladas en la producción de carne, la conservación de 100,000 hectáreas forestales e ingresos adicionales para los productores locales por más de 250 millones de pesos. Aún ahora, no cuenta con un sistema de monitoreo en marcha para verificar avances en estas metas.

Como parte de estándares mínimos de evaluaciones *ex post*, tanto el CONACYT como la propia universidad, podrían, con base en lo establecido en la teoría del cambio de los proyectos, solicitar pruebas de interacciones productivas con otros actores. Ejemplo de esto podrían ser indicadores que reflejen la influencia que el conocimiento ha tenido, tales como el desarrollo de habilidades, el comienzo de acciones o el establecimiento de consensos (Kumar *et al.*, 2018). Por otro lado, podrían emplearse indicadores que reflejen el reconocimiento obtenido, mediado a través de financiamiento adicional gestionado o la invitación a participar en un espacio de toma de decisiones (Bensing *et al.*, 2013).

Es interesante constatar que varios indicadores para medir el cumplimiento de la demanda original del Fondo Mixto tenían esta visión (número de redes científicas en las que participa el centro, número de programas transmitidos en radio y televisión, convenios de colaboración con otras organizaciones, manuales de buenas prácticas con calendarios de capacitación, estrategias para atender intereses de productores). Sin embargo, la mayor parte no se utilizó en la evaluación final del proyecto, además de que la mera entrega de un documento con procedimientos organizacionales no garantiza su uso en la operación cotidiana.

En este mismo sentido, la investigación mostró una brecha en la implementación de evaluaciones orientadas al aprendizaje, las cuales se plantearon en el papel para retroalimentar el plan de desarrollo del CPIA, pero no se llevaron a cabo. Para trascender las buenas intenciones, se requiere que el procedimiento se incorpore como parte del modelo de gestión del CPIA, que los investigadores conciban las evaluaciones como una herramienta valiosa y que en el equipo exista personal con la función y la capacidad de facilitar evaluaciones formativas.

Al interior de la universidad, para vincular la medición del desempeño con la orientación a resultados se requiere del apoyo de las autoridades para transformar las reglas del juego. Por un lado, al asignar recursos para desarrollar capacidades de evaluación en equipos interdisciplinarios de investigación. Por el otro, al modificar el sistema de incentivos, de forma que los investigadores sean recompensados por evaluar la efectividad de sus proyectos.

A continuación, se presenta un ejemplo del tipo de mecanismos que se podrían emplear. Para ello, se retoma el caso de un proyecto actual del CPIA, el cual tiene el objetivo de recopilar y promover la diversidad biocultural de la reserva a través de una estrategia de conservación

ex situ. El proyecto, financiado por CONACYT, reunirá en el CPIA varias colecciones de especies arbóreas, no maderables y de sistemas productivos tradicionales.

Después de revisar el diseño del proyecto, se identificó que la lógica causal de los cambios esperados quedó en una caja negra, pues no se mencionaba con claridad cómo las estrategias de difusión servirían para interactuar con actores específicos y cómo en consecuencia se generarían resultados medibles. La descripción de resultados del proyecto finaliza con una serie de productos entregables (Figura 1).

Figura 1. Representación de la lógica causal del proyecto de conservación *ex situ*



Fuente: Elaboración propia

En concreto, se propone integrar **rutas de generación de impactos** como un estándar de calidad en el diseño de proyectos de investigación aplicada. Esto permitiría fortalecer las cadenas de resultados y transparentar el proceso de cambio esperado. Se propone la integración de los siguientes elementos, retomados de varias metodologías para evaluar la intermediación de conocimientos (Earl *et al.*, 2001; Alvarez *et al.*, 2010; Van der Meulen *et al.*, 2010; Spaapen *et al.*, 2011; Kumar *et al.*, 2018; Tilley *et al.*, 2018; Glover *et al.*, 2019).

1. Especificar los conocimientos que se espera transferir, los cuales pueden ser materiales (plántulas de especies nativas, infraestructura de irrigación) y/o técnicas (pastoreo racional, producción de bioinsumos). Enunciarlos permite identificar a los actores involucrados y evidenciar los supuestos que están detrás de las intenciones de cambio (los ganaderos están interesados en sistemas silvopastoriles, siempre y cuando incrementen la disponibilidad de forraje). En el cuadro 6 se ejemplifica su uso en el contexto del proyecto de conservación *ex situ*.

Cuadro 6. Ruta de generación de impacto del proyecto de conservación *ex situ* (I).

Conocimientos	Actores	Supuestos
<ul style="list-style-type: none"> • Recopilación de conocimientos tradicionales. • Plántulas de especies nativas. • Técnicas de manejo para instalar traspatios agroforestales. 	Unidades domésticas de comunidades de la región	Algunas viviendas tienen solares abandonados con el potencial de integrar especies nativas en traspatios agroforestales.
<ul style="list-style-type: none"> • Plántulas de especies nativas. • Elementos para diseñar reglamentos y programas de fomento de especies nativas. 	Ayuntamientos de la reserva	Los ayuntamientos pueden destinar recursos para programas de subsidio que fomenten el uso de especies nativas.

Fuente: Elaboración propia

2. Especificar las estrategias de interacción con los actores involucrados, así como los efectos esperados al interactuar con el conocimiento. Al hacerlo, hay que clarificar el tipo de cambio, el grado de dificultad y el indicador con el que se mediría. Esto implica, por necesidad, preparar instrumentos para identificar los cambios que se pretende generar. En el cuadro 7 se ejemplifica su uso en el contexto del proyecto de conservación *ex situ*.

Cuadro 7. Ruta de generación de impacto del proyecto de conservación *ex situ* (II)

Estrategias de interacción	Tipo de cambio	Cambio esperado	Indicador	Grado de dificultad
Taller de transferencia tecnológica	Desarrollo de habilidades	Los participantes planifican lo que desean para su solar y aprenden cómo mantener un traspatio agroforestal.	% de participantes que, 3 meses después de los talleres, formalizaron la solicitud de especies nativas para establecerlas en su traspatio.	Se espera observar
Asistencia de pares	Comienzo de acciones	Los participantes instalan un traspatio agroforestal en su vivienda.	Número de traspatios agroforestales en funcionamiento, dos años después de su instalación.	Se desea observar
Evaluación del retorno social de la inversión	Generación de beneficios	Los participantes obtienen beneficios económicos y sociales por rehabilitar sus solares.	Monetarización de los productos obtenidos del traspatio agroforestal (forraje, leña, plantas medicinales, hortalizas, material de construcción)	Sería excelente observar
Diálogos multi-actor	Establecimiento de consensos	Existe respaldo para elaborar un reglamento intermunicipal de forestación.	Reglamento elaborado y aprobado por los cabildos.	Se desea observar

Fuente: Elaboración propia

En conjunto, el procedimiento propuesto sentaría las bases de:

- Evaluaciones participativas *ex ante*, a través de las cuales se coloque en el centro la capacidad de agencia de la población objetivo en el diseño de las investigaciones.
- El monitoreo de la dinámica del proceso, a medida que se desarrollan las intervenciones.
- El uso de evaluaciones como una herramienta de mejora continua, pues el aprendizaje obtenido serviría para retroalimentar los proyectos.

Sin embargo, el uso de esta propuesta, como parte de los procedimientos del CPIA, requiere cambios organizacionales más amplios. Si la universidad a la que pertenece no ajusta su sistema de evaluación, los investigadores invertirían tiempo en actividades no valoradas por su organización, ralentizando el avance en su carrera profesional. De esta manera, un análisis costo/beneficio por parte de los investigadores podría obstaculizar su adopción.

Por este motivo, sería necesario que la universidad **vinculara la valoración del desempeño con indicadores de resultado**. En específico, se propone que equipos interdisciplinarios de investigadores sean evaluados con base en indicadores seleccionados por ellos mismos, los cuales, a su parecer, reflejen tanto la relevancia de sus proyectos como el proceso de aprendizaje que han tenido.

Dado que el diálogo franco sobre los indicadores de desempeño es lo que convierte la información en conocimiento (Serrat, 2010), el medio de rendición de cuentas serían autoevaluaciones donde los equipos de investigación sintetizen el cambio que han conseguido, valoren la efectividad de los mecanismos de intercambio, identifiquen desviaciones y acuerden acciones correctivas para obtener mejores resultados. La intención del procedimiento es estimular el trabajo colaborativo e interdisciplinario orientado hacia la obtención de resultados.

Actualmente, los investigadores de la universidad comprometen productos académicos. La propuesta de mejora es complementar esa métrica con indicadores que indiquen que sus interacciones han sido productivas (Cuadro 8). Para integrarla dentro de sus procedimientos, la universidad podría emplear distintos mecanismos, desde convocatorias que condicionen el financiamiento a su cumplimiento, hasta el rediseño del reglamento de ingreso y promoción del personal académico de la universidad, de forma que se integren indicadores que recompensen la trayectoria de académicos orientados a la transformación de su entorno.

Cuadro 8. Propuesta de indicadores para valorar el desempeño de investigadores

	INVESTIGACION	RELEVANCIA
Demstrar la generación de un producto	Artículo en revista indizada	Evidencia de talleres, programas de radio, videos, notas periodísticas
Demstrar el uso del conocimiento	Índice h	Indicadores seleccionados que reflejan cambios en la toma de decisiones de actores involucrados
Demstrar el reconocimiento	Financiamiento, premios	Financiamiento, premios, invitación a un comité, altmetrics

Modificado a partir de Bensing *et al.* (2013) y Hicks *et al.* (2015).

Conclusiones

Las evaluaciones asociadas al CPIA tuvieron tres propósitos: verificar la relevancia de los proyectos antes de su aprobación, rendir cuentas sobre los resultados inmediatos, obtenidos con los proyectos, y tomar decisiones laborales sobre el personal académico. No se han realizado evaluaciones para estimar los cambios generados, ni para retroalimentar los proyectos con la finalidad de mejorarlos.

Se considera necesario reformar el sistema de evaluación al interior de la universidad, de tal manera que mida el valor público generado mediante los proyectos de investigación agrícola desarrollados en el CPIA. Para empezar, hay que cambiar la forma en la que se concibe la calidad de las investigaciones. El paradigma actual asume que la productividad académica es algo inherentemente bueno y, por tanto, una medida suficiente para que la ciencia rinda cuentas. Lamentablemente, si la valoración del desempeño de los investigadores no promueve que se aboquen a la consecución de valor público con sus proyectos, se corre el riesgo de que algunos de ellos prioricen acciones que «dan puntos», más allá del impacto social que puedan tener sus investigaciones (Suárez y Muñoz, 2016).

En cambio, un esquema orientado al impacto promovería que los investigadores, más allá del cumplimiento de los estándares metodológicos de su disciplina, formen equipos interdisciplinarios e inviertan en actividades de intermediación del conocimiento (Bensing *et al.*, 2013). Esta perspectiva ampliada de la rendición de cuentas requiere que el investigador aporte pruebas de que está comprometido con los objetivos que desea conseguir (Ebrahim, 2003; Guijt, 2010). El artículo presentó una propuesta procedimental para hacer realidad esta visión. Se considera que la universidad estudiada tiene el suficiente margen de maniobra para integrar los cambios sugeridos.

Capítulo seis

¿Medimos lo que es valioso o valoramos lo que es medible? Comparación interorganizacional del uso de evaluaciones en una reserva de la biosfera

Do we measure what is valuable or value what is measurable? Interorganizational Comparison of Evaluation Use in a Biosphere Reserve

Eduardo Luna-Sánchez, Margaret Skutsch

Revista: **Gestión y ambiente**; Año de publicación: **2020**

Liga de descarga:

<https://revistas.unal.edu.co/index.php/gestion/article/view/79714>

¿Medimos lo que es valioso o valoramos lo que es medible? Comparación interorganizacional del uso de evaluaciones en una reserva de la biósfera

Eduardo Luna Sánchez y Margaret Skutsch

Resumen

La investigación compara el grado en que cuatro organizaciones usan evaluaciones para identificar la obtención de resultados y mejorar su práctica. Las organizaciones estudiadas (una organización de la sociedad civil (OSC), una dependencia federal, una universidad pública y una administración municipal) llevan a cabo proyectos de manejo de recursos naturales en una reserva de la biósfera terrestre de México. En el marco de un estudio de casos, se combinó el análisis de documentos institucionales con entrevistas semi-estructuradas para acceder a la información. La OSC tiene el sistema de evaluación más adecuado para identificar el cumplimiento de objetivos de conservación y desarrollo. En contraste, la administración municipal carece por completo de procedimientos evaluativos.

Conceptos clave: Aprendizaje organizacional; Cultura organizacional; Rendición de cuentas; Desarrollo de capacidades en evaluación

Abstract

The research compares the degree to which four organizations use evaluations to identify outcomes and to improve their practice. The studied organizations (a nonprofit organization (NGO), a federal agency, a local government and a public university) implement natural resource management projects in a terrestrial biosphere reserve of Mexico. Framed in a case study, an analysis of institutional documents was combined with semi-structured interviews to gain access to the information. The NGO has the most appropriate evaluation system to identify compliance with conservation and development objectives. In contrast, the local government completely lacks evaluative procedures.

Keywords: Organizational learning; Organizational culture; Accountability; Evaluation Capacity Building

Introducción

La gestión de Reservas de la Biósfera (RB) tiene dos objetivos centrales: garantizar la protección del patrimonio natural y el bienestar de los pobladores locales (UNESCO, 2013). Por este motivo, en ellas coexisten iniciativas de conservación y desarrollo, no siempre coordinadas entre sí, como el ecoturismo, la producción agropecuaria o el manejo forestal (Baird, 2014).

Es complejo obtener resultados sociales y ambientales de manera simultánea, por lo que se requiere respaldar con evidencia si las buenas intenciones han tenido éxito. Esto ha impulsado la producción de evaluaciones, dada su utilidad para identificar cambios en variables sociales, económicas y ambientales (Ferraro y Pressey, 2015).

En contextos complejos, como el manejo de RB, se espera que sirvan para rendir cuentas y orientar la acción futura (Margoluis *et al.*, 2013; Pressey *et al.*, 2015). Por ello, abundan las metodologías de evaluación orientadas al manejo adaptativo de los proyectos (Pomeroy *et al.*, 2004; Hockings *et al.*, 2015; Stolton y Dudley, 2016).

Sin embargo, aunque se espera aprender a partir de ellas, diversas investigaciones sobre el uso de las evaluaciones han demostrado que con frecuencia son irrelevantes o afectan negativamente el comportamiento organizacional (Ebrahim, 2005; Guijt, 2008). Algunas barreras referidas son la falta de recursos y el temor a las consecuencias en caso de reportar resultados negativos (Craigie *et al.*, 2015; McKinnon *et al.*, 2015).

En general, prevalece el uso de evaluaciones de desempeño que se limitan a medir el cumplimiento de las actividades comprometidas (Stem *et al.*, 2005). Esto es insuficiente, pues los resultados inmediatos son un mal predictor de la generación de beneficios de conservación y desarrollo (Kapos *et al.*, 2009).

Proponer a priori ciertos indicadores o metodologías podría pasar por alto los dilemas y restricciones a los que se enfrenta el personal operativo de distintas organizaciones. Por estos motivos, incorporar las evaluaciones como una herramienta de aprendizaje requiere una buena comprensión del funcionamiento de las organizaciones, así como de los arreglos institucionales bajo los que operan (Guijt, 2008).

En México, el estudio sobre el uso de evaluaciones se ha concentrado en torno de la política de evaluación del gobierno federal (Cejudo y Maldonado, 2011; Cardozo, 2015). En el contexto de RB, las investigaciones se han enfocado en el análisis de evaluaciones externas de programas federales de subsidio (Orta *et al.*, 2013; Sotelo *et al.*, 2015). Si bien se han llevado a cabo evaluaciones en la escala de RB (Guerra y Ochoa, 2008; Ramos *et al.*, 2018), no se ha estudiado su uso entre organizaciones directamente involucradas en la operación de proyectos. Conocer el fenómeno es relevante, pues deberían ser las principales usuarias de evaluaciones que les sirvan para manejar los proyectos de manera adaptativa.

Un análisis interorganizacional del uso de evaluaciones serviría para identificar áreas de sinergia que complementen esfuerzos evaluativos aislados. Esto es especialmente útil en una RB, donde diferentes organizaciones trabajan con objetivos similares. El proceso de

comunicación resultante facilitaría la coordinación y el anidamiento de esfuerzos para medir y explicar el progreso hacia metas comunes de conservación y desarrollo.

Este artículo presenta un análisis comparativo de la forma en la que distintos tipos de organización hacen y usan evaluaciones en una RB terrestre de México. Todas ellas (una Administración Municipal (AM), una dependencia del gobierno federal, una OSC y una universidad pública) llevan a cabo proyectos vinculados con el manejo de recursos naturales.

Es importante resaltar que la investigación no busca valorar qué tan competentes o efectivas son las organizaciones. En cambio, las preguntas que se busca responder son:

-¿Cómo y para qué hacen evaluaciones?

-¿Qué factores promueven o inhiben su uso como una herramienta de aprendizaje organizacional?

La información está organizada de la siguiente manera. Primero, se describe el marco conceptual a partir del cual se estructura el análisis y después se detalla la metodología utilizada para responder las preguntas de investigación. En la sección de resultados se describen las características de las organizaciones y se compara el propósito de las evaluaciones que realizan. En la discusión, se contrastan los hallazgos con la literatura sobre el tema, al analizar los factores que inciden en el uso de las evaluaciones como una herramienta de aprendizaje. Finalmente, en la conclusión se sintetizan los mensajes centrales del artículo.

El uso de evaluaciones como objeto de estudio

Las evaluaciones pueden realizarse para diferentes propósitos, como la rendición de cuentas, la planeación de proyectos o el aprendizaje organizacional. Si bien, se asume que las evaluaciones sirven simultáneamente para todos esos fines, históricamente se ha privilegiado la valoración del desempeño organizacional para prevenir el uso inadecuado de fondos públicos (Bovens *et al.*, 2008; Cousins *et al.*, 2015). Normalmente, el objetivo de ese tipo de evaluaciones es demostrar transparencia financiera y administrativa, al aportar pruebas de la implementación de los planes acordados y reportes de los gastos realizados (Ebrahim, 2005; Guijt, 2008).

Como herramienta de aprendizaje organizacional, las evaluaciones aportan información para que las organizaciones adapten su toma de decisiones y consigan mejores resultados (Chelmsky, 1997). De manera *ex ante*, las evaluaciones mejoran el diseño de los proyectos al corregir errores en la lógica de una intervención o al identificar el riesgo de generar impactos negativos. Al hacerse de manera simultánea a la implementación de un proyecto, las evaluaciones sirven para hacer adecuaciones sobre la marcha. De forma *ex post*, se espera que las evaluaciones de impacto identifiquen cambios ligados a las actividades y productos de un proyecto. Esto permite rendir cuentas sobre el valor público generado, y a la vez, retroalimentar procesos de planeación a través de la identificación de fallas (Richards, 2012).

Además, se han propuesto metodologías que, más allá de proyectos, evalúan la efectividad organizacional. La intención es identificar áreas de oportunidad y proponer procedimientos

que mejoren su funcionamiento. En particular, en el campo de la conservación se cuenta con metodologías como el Índice de Implementación y Gestión (OLCEFS⁶⁶, 2015), o la Herramienta de Seguimiento de la Efectividad del Manejo (Stolton y Dudley, 2016).

El malestar generado por el bajo impacto de las evaluaciones promovió el estudio de la forma en la que son utilizadas (Leviton y Hughes, 1981). El uso de las evaluaciones ha sido explicado a través de factores relacionados con el cómo, por quién y para quién se conduce una evaluación, los cuales son definidos y alterados por el contexto organizacional e institucional (Taut, 2007) (Figura 1).

En primera instancia, se han señalado variables relacionadas con la calidad de las evaluaciones. En esta escala, basada en aspectos metodológicos, el uso se explica en función de la disponibilidad de información confiable cuando es requerida para tomar decisiones (Alkin, 1985). Esto implica que los proyectos son evaluables, pues su lógica interna permite asumir que las actividades provocarán los cambios deseados. Además, conducir evaluaciones requiere conocimientos técnicos que garanticen la calidad de la información generada⁶⁷ (Cousins *et al.*, 2014).

Figura 1. Factores que afectan el uso de las evaluaciones.



Fuente: Elaboración propia

En el siguiente nivel, influyendo sobre la calidad, se encuentran variables organizacionales. Las organizaciones son el principal usuario de las evaluaciones; por ello, es importante entender si están preparadas para hacerlas y retomar sus resultados. La práctica evaluativa requiere el desarrollo de puestos de trabajo, sistemas y planes que integren las evaluaciones en la forma en que la organización cumple con sus objetivos (Preskill y Boyle, 2008, p. 444). Además, son fundamentales la disponibilidad de recursos (tiempo y dinero) y un ambiente organizacional que favorezca el reporte de fallas y la transferencia de lo aprendido al trabajo cotidiano (Cousins *et al.*, 2014).

En última instancia, las organizaciones operan guiadas por arreglos institucionales que influyen en su comportamiento. De esta manera, por un mandato o por su dependencia hacia recursos externos, sus evaluaciones pueden seguir parámetros y propósitos no definidos por

⁶⁶ Organización Latinoamericana y del Caribe de Entidades Fiscalizadoras Superiores.

⁶⁷ Por ejemplo, elegir la metodología en función del propósito, seleccionar indicadores pertinentes, o saber cómo recolectar datos.

ellas mismas (Guijt, 2008). En cualquier caso, es importante identificar el margen de maniobra de las organizaciones para adoptar o no los procedimientos, así como para desarrollar esquemas de evaluación siguiendo sus propios propósitos.

Metodología

Dada la diversidad de escenarios e interacciones involucradas, se ha reconocido que es muy complejo generalizar y jerarquizar las características que más se relacionan con el uso de evaluaciones (Weiss, 1998; Johnson *et al.*, 2009).

Por ello, se considera pertinente el estudio de casos para estudiar el uso de las evaluaciones, pues el método permite vincular datos empíricos con proposiciones teóricas, con lo que se prueba su validez al estudiar un fenómeno en su propio contexto (Yin, 2009).

De esta manera, en el marco de un estudio comparativo de casos, se combinó el análisis de documentos institucionales con entrevistas semi-estructuradas para acceder a la información. La recopilación de documentos relacionados con las evaluaciones se llevó a cabo a través de una exploración de la página web institucional de cada una de las organizaciones, solicitudes de acceso a la información y el uso de motores de búsqueda en internet.

Se hicieron 29 entrevistas semi-estructuradas a personal de las organizaciones relacionado con procedimientos de evaluación (Cuadro 1). Éstas permitieron recolectar información sobre la puesta en práctica de las evaluaciones, entremezclada con la opinión y reflexiones de los entrevistados.

La información de las evaluaciones es especialmente sensible, pues la posibilidad de recibir una evaluación negativa genera ansiedad (Donaldson *et al.*, 2002). A nivel organizacional, esto se traduce en el temor de perder credibilidad y de recibir sanciones en caso de no cumplir con expectativas de desempeño (McKinnon *et al.*, 2015). Por ello, con la finalidad de evitar la percepción de un riesgo potencial, se tomó la decisión de garantizar el anonimato de las organizaciones que participaron en la investigación. Como consecuencia, no se brinda información específica sobre la RB en la que trabajan.

Cuadro 1. Fuentes de información de la investigación.

Organización	Número de entrevistas	Número de documentos revisados
Dependencia federal	11	35
Universidad Pública	10	22
Administración municipal	4	17
Organización de la Sociedad Civil	4	24

Fuente: Elaboración propia

El contexto organizacional del estudio

La RB en la que operan las organizaciones cuenta con cerca de 25,000 hectáreas de zona núcleo, las cuales protegen el hábitat de distintas especies, como el emblemático jaguar. En ella, la conservación de la biodiversidad enfrenta una serie de amenazas, como la ganadería extensiva en áreas forestales, la falta de regulación de actividades turísticas, o la contaminación de agua por descargas residuales y actividades mineras. La población total, dispersa en más de 600 localidades, se acerca a los 100,000 habitantes. A continuación, se presentan las características de las organizaciones estudiadas y se enlistan los objetivos de sus proyectos (Cuadro 2).

Dependencia federal. La dependencia federal es la Comisión Nacional de Áreas Naturales Protegidas (CONANP), entidad responsable de conservar la biodiversidad en los ecosistemas más representativos de México. A escala local, la CONANP se apoya en direcciones técnicas para administrar Áreas Naturales Protegidas (ANP). En ellas, equipos de guardaparques llevan a cabo acciones de conservación siguiendo un programa de manejo (CONANP, 2014). En la RB, la dirección de ANP se conforma por diez guardaparques, los cuales cuentan con un perfil profesional relacionado con el manejo de recursos naturales (biología, ingeniería forestal, veterinaria) y, en menor medida, con áreas administrativas (administración, turismo).

Universidad Pública Estatal. Las Universidades Públicas Estatales son instituciones de educación superior que reciben financiamiento público para llevar a cabo las funciones de docencia, generación y aplicación del conocimiento. Debido a su alta productividad científica, la universidad estudiada es considerada una de las mejores del país. La universidad cuenta con la mayor cantidad de carreras en el estado, las cuales oferta no sólo en la ciudad capital, sino en diferentes municipios. Esto es así, pues una de sus políticas ha sido la promoción del desarrollo a través del establecimiento de campus regionales. Dentro de la RB, la universidad construyó un centro de investigación con el objetivo de atender problemáticas locales. A través de él, se forman recursos humanos especializados y se promueve la transferencia de tecnología para los diferentes procesos ligados al manejo de recursos naturales. El centro de investigación es multidisciplinario, en él colaboran diez profesores-investigadores con perfiles complementarios (ciencias biológicas, agropecuarias y sociales).

Administración Municipal. En México, los municipios son la base de la división territorial, política y administrativa. Éstos gozan de autonomía, por lo que pueden definir su estructura organizacional y su normatividad interna. Para su gobierno, la ciudadanía elige a un ayuntamiento por periodos trianuales (CONEVAL, 2016). El municipio cuya administración se estudió está inserto en su totalidad dentro de la RB. Debido a su relación con evaluaciones y con proyectos ligados al manejo de recursos naturales, se analizaron los procedimientos utilizados por la Contraloría Municipal, la Dirección de Desarrollo Agropecuario, y la Coordinación de Ecología.

Organización de la Sociedad Civil. La Organización de la Sociedad Civil (OSC) tiene el objetivo central de conservar la biodiversidad de la RB. La OSC ha resaltado por su

capacidad de gestión financiera, contando a lo largo de tres décadas con una larga lista de donantes nacionales e internacionales. A nivel regional, la OSC encabeza una alianza de cinco organizaciones que comparten la misma dirigencia. Su finalidad ha sido abarcar diferentes áreas de la conservación y el desarrollo, como el ecoturismo, prácticas agropecuarias sostenibles o la gestión de pagos por servicios ambientales para propietarios de terrenos forestales.

Cuadro 2. Programas y proyectos emprendidos por las distintas organizaciones.

Organización	Personal⁶⁸	Objetivos de programas y proyectos
Dependencia Federal	10	<ul style="list-style-type: none"> • Conservación de especies en riesgo • Rehabilitación forestal • Educación ambiental • Monitoreo biológico • Promoción de proyectos productivos sostenibles • Regulación de actividades turísticas • Control de incendios • Conservación del maíz criollo
Universidad Pública Estatal	13	<ul style="list-style-type: none"> • Transferencia de tecnología: <ul style="list-style-type: none"> - Producción de ovinos - Nutrición animal - Sistemas silvopastoriles - Plantaciones bioenergéticas - Agricultura protegida - Producción de bioinsumos - Ecotecnias - Reforestación con especies nativas - Banca comunitaria
Administración municipal	7	<ul style="list-style-type: none"> • Subsidios para adquirir insumos e infraestructura que fortalezca la producción agropecuaria • Promoción de la producción de traspatio • Siembra de árboles frutales • Instalación de estufas ahorradoras de leña
Organización de la Sociedad Civil	18	<ul style="list-style-type: none"> • Compra de tierras para consolidar una red de reservas privadas • Mecanismos de conservación basados en mercado • Manejo regenerativo de suelos • Monitoreo biológico • Fortalecimiento de microoperadores turísticos • Fortalecimiento de proyectos productivos sostenibles • Educación ambiental • Manejo de residuos sólidos en comunidades

Fuente: Elaboración propia

⁶⁸ En la cifra se coloca el personal operativo y administrativo vinculado con los proyectos referidos.

Comparación interorganizacional del uso de evaluaciones

En este apartado se describen las evaluaciones utilizadas por las organizaciones. Los resultados muestran diferencias en los tipos de evaluación empleados (Cuadro 3). Se distinguen tres propósitos asociados a las evaluaciones: rendición de cuentas (RC), aprendizaje organizacional (AO), y gestión financiera (GF).

Cuadro 3. Tipos de evaluación utilizados por las organizaciones.

TIPO DE EVALUACIÓN	ORGANIZACIÓN	PROPÓSITO DE LA EVALUACIÓN		
		RC	AO	GF
Evaluación de desempeño	Dependencia federal	RC	AO	
	OSC	RC	AO	
	Universidad Pública	RC		
Evaluación <i>ex ante</i>	OSC	AO	GF	
Evaluación de la efectividad organizacional	Dependencia federal	RC	AO	GF
	Administración municipal	AO		
Evaluación de impacto	Dependencia federal	AO		
	OSC	RC	AO	GF

Fuente: Elaboración propia

Evaluación del desempeño basado en la generación de resultados inmediatos.

Este tipo de evaluación es la más recurrente entre las organizaciones estudiadas. En el caso de la dependencia federal, el equipo de guardaparques elabora un Programa Operativo Anual (POA). En él, siguiendo las directrices del programa de manejo de la RB, establecen las actividades que realizarán para lograr sus objetivos. Aun así, su POA debe contemplar estrategias y unidades de medida establecidas desde la administración central, con la finalidad de estandarizar la información reportada por cada Dirección de ANP del país. Esto le permite a la CONANP alimentar indicadores nacionales que miden la obtención de resultados inmediatos, como la instalación de consejos ciudadanos o el número de comunidades atendidas.

Como mecanismo de evaluación de desempeño, el POA se utiliza para monitorear trimestralmente el cumplimiento de las actividades comprometidas por los guardaparques. Anualmente, el personal directivo de la dependencia utiliza las evaluaciones para retroalimentarlos y tomar decisiones presupuestales. En la escala local, la elaboración anual del POA es un momento de reflexión que permite a los guardaparques hacer ajustes al recuperar el aprendizaje acumulado con la implementación de los proyectos.

Aun así, dado el temor a recibir malas valoraciones, algunos guardaparques identificaron el riesgo de privilegiar resultados seguros, pero superficiales y efímeros. Esto, señalaron, se ve exacerbado por la dinámica de la administración pública, la cual promueve proyectos ligados al año fiscal a los que después no se les da seguimiento.

En el caso de la AM, a pesar de que su reglamento interior solicita que cada área cuente con sistemas de evaluación, la organización no cuenta con procedimientos para identificar la obtención de resultados inmediatos. Ni el plan municipal de desarrollo ni los POA de las áreas planteaban indicadores para ese fin. La rendición de cuentas, mediada a través de informes de gobierno, se limita a señalar el monto ejercido y el número de beneficiarios ligados a cada proyecto.

La OSC estudiada ha interactuado con múltiples evaluaciones de desempeño de proyectos financiados por donantes nacionales e internacionales. Por ello, existen variaciones importantes en las características de cada evaluación. Dependiendo del donante, los mecanismos utilizados incluían: auditorías a la ejecución del presupuesto, reportes del cumplimiento de actividades comprometidas, evaluaciones parciales para identificar y corregir fallas en la implementación, o evaluaciones finales para demostrar la obtención de resultados.

Las principales diferencias se encuentran en proyectos multianuales financiados por donantes internacionales como el Banco Interamericano de Desarrollo (BID) o el Fondo Global para el Medio Ambiente (GEF), los cuales solicitan el establecimiento de líneas de base y sistemas de monitoreo para medir los indicadores con los que se valorará el desempeño. Este tipo de proyectos, además de reportes periódicos para verificar la correcta ejecución del presupuesto y el cumplimiento de los compromisos establecidos en los convenios de colaboración, recurren a evaluaciones intermedias para proponer modificaciones sobre la marcha.

En el caso del centro de investigación de la universidad, las evaluaciones de desempeño tienen diferentes propósitos. Por un lado, se encuentran procedimientos para evaluar el desempeño de proyectos con financiamiento externo, los cuales siguen requisitos dictados por los donantes, y por el otro, procedimientos para evaluar el desempeño de los investigadores que son establecidos por la propia universidad.

Respecto del primer tipo de evaluación, el centro de investigaciones surgió a partir de una convocatoria de Fondos Mixtos del CONACyT⁶⁹, con el objetivo de transferir conocimiento científico para mejorar los procesos productivos del sector rural en la RB. La evaluación externa del proyecto valoró la entrega de diversos productos que tenían la intención de preparar la actuación del centro. De esa manera, entre los entregables estaba un plan estratégico, un calendario de capacitaciones y la generación de ocho publicaciones

⁶⁹ El Consejo Nacional de Ciencia y Tecnología (CONACyT) es un organismo público descentralizado del gobierno federal mexicano que se encarga de promover el avance de la investigación científica.

científicas. Es importante señalar que, si bien el plan estratégico establecía indicadores para verificar la obtención de resultados, a la fecha no se ha realizado una evaluación de los resultados obtenidos.

Respecto del segundo tipo de evaluación, la universidad utiliza un sistema cuantitativo de evaluación académica para normar el proceso a través del cual los investigadores pueden ganar plazas de tiempo completo, recibir estímulos y obtener promociones laborales. Su objetivo es promover en los investigadores un comportamiento alineado a los indicadores de productividad académica con los que se evalúa el desempeño global de la universidad. Por ello, destacan por su valor las publicaciones, ponencias en congresos, registros de propiedad intelectual o tesis.

Evaluación ex ante

La OSC es la única que ha interactuado con evaluaciones *ex ante*. Las evaluaciones han sido parte del proceso de gestión de proyectos multianuales o una condición necesaria para acceder al mercado voluntario de carbono. El propósito central de estas evaluaciones ha sido garantizar que los proyectos financiados cumplen con estándares de calidad en su diseño⁷⁰.

Algunos elementos revisados por este tipo de evaluaciones fueron:

-El valor de verdad de la secuencia lógica del proyecto (*Actividad-Output-Outcome*), así como los supuestos y los riesgos asociados. Para ello, requirieron el uso de modelos lógicos, como teorías del cambio.

-Planes de monitoreo y evaluación “creíbles” y “rigurosos” que establezcan líneas de base, indicadores de impacto y la metodología que se seguirá. También, con la finalidad de estimar la adicionalidad ofrecida tanto en la captura de carbono como en el bienestar de la población objetivo, se ha requerido la descripción de escenarios sociales de referencia en ausencia de proyecto.

-Evaluaciones de los impactos potenciales sobre grupos vulnerables, con la finalidad de garantizar que el proyecto no les haría daño.

Evaluación de la efectividad organizacional

Este tipo de evaluaciones fueron llevadas a cabo por las dos organizaciones gubernamentales. La CONANP utiliza diferentes metodologías para medir la efectividad del manejo que realiza en las ANP. Este tipo de evaluaciones no valoran en sentido estricto la efectividad de las acciones, sino la presencia de elementos que se consideran imprescindibles en un buen ciclo de gestión de proyectos de conservación (p. ej. actividades de monitoreo que retroalimenten las decisiones de los guardaparques, la existencia de un documento de planeación estratégica o la suficiencia de recursos para la operación).

En la RB, el equipo de guardaparques ha calculado el índice de implementación y gestión (OLCEFS, 2015), conocido también como INDIMAPA, para determinar el grado en el que la dirección técnica posee las condiciones institucionales, normativas y operativas que son

⁷⁰ Como los estándares de Clima, Comunidad y Biodiversidad (CCBA, 2013).

necesarias para poder ser efectiva. Además, actualmente se encuentra en proceso una evaluación utilizando la metodología i-efectividad, desarrollada por la CONANP a partir de otros instrumentos de evaluación⁷¹ para calcular un índice de efectividad en el manejo de las ANP federales de México.

A nivel nacional, las evaluaciones le permiten a la CONANP rendir cuentas sobre el cumplimiento de la Meta 11 de Aichi⁷², y con ello, facilitar la gestión de recursos para cubrir parte de la brecha financiera de las ANP en México. A nivel local, el propósito es retroalimentar al equipo de guardaparques para que implemente las recomendaciones de mejora. Desde la perspectiva del director del ANP, la información de INDIMAPA ha sido útil para que la DES haga un diagnóstico global de la situación de las ANP en el país, pero no para tomar decisiones directamente en el área. La implementación de la evaluación i-efectividad se encontraba en marcha cuando se hizo la investigación, por lo que no se cuenta con datos sobre su influencia.

Por su parte, la AM decidió de manera voluntaria participar en el programa “Agenda para el Desarrollo Municipal”, impulsado por el Instituto Nacional para el Federalismo y el Desarrollo Municipal (INAFED). A través de él, recibió en 2016 servicios de consultoría para operar un sistema de evaluación de la gestión pública municipal. El propósito del sistema es identificar áreas de mejora y establecer plazos para que el municipio cumpla con el desempeño esperado en sus responsabilidades constitucionales.

En materia de evaluación, algunos de los elementos valorados son la existencia de instrumentos de planeación que cuenten con una ficha técnica de indicadores de seguimiento y cumplimiento, y la existencia de una instancia responsable de evaluar el cumplimiento de los instrumentos de planeación. Con respecto del segundo punto, la contraloría municipal es la encargada de evaluar la gestión pública, sin embargo, dado que no existen indicadores, su revisión se limita a verificar que cada área tenga su POA.

Evaluación de impacto

Las organizaciones que tienen procedimientos relacionados con la evaluación de impactos son la CONANP y la OSC. En el caso de la CONANP, por ley los programas de manejo de las diferentes ANP del país deben ser revisados cada cinco años con el objeto de evaluar su efectividad y proponer modificaciones.

Para cumplir con este mandato, la CONANP desarrolló una cédula de evaluación de la efectividad con la que los guardaparques evalúan el cumplimiento de los objetivos trazados en los diferentes componentes del programa y, en caso de ser necesario, explican las razones por las que hay que replantear las estrategias y acciones. Para hacer el análisis, es necesario que exista un sistema de “seguimiento y monitoreo de los proyectos” que permita evaluar los avances anualmente.

⁷¹ Principalmente de la Herramienta de Seguimiento de la Efectividad del Manejo (METT) (Stolton y Dudley, 2016).

⁷² Esta meta establece que para el año 2020, el 17 % de la parte terrestre de México, así como el 10 % de su superficie marina estarán bajo un régimen de conservación, y que esa superficie será administrada de manera efectiva.

En la reserva, el plan de manejo vigente tiene más de 15 años de antigüedad. De 2014 a la fecha el equipo de guardaparques ha trabajado en su evaluación con miras a que sea actualizado. Existen dos obstáculos para que, como tal, este procedimiento evalúe el impacto que ha tenido la dependencia en la reserva. En primer lugar, por falta de recursos, los guardaparques no dan seguimiento formal a los proyectos por lo que no poseen información sobre sus efectos. En segundo lugar, la mayor parte de los indicadores del actual programa de manejo sólo estiman la obtención de resultados inmediatos (p. ej. número de investigaciones científicas o número de proyectos productivos sustentables). Como resultado, la evaluación elaborada por los guardaparques se concentró en el cálculo de porcentajes de implementación al comparar las acciones realizadas con las acciones originalmente previstas en el programa de manejo.

Independientemente del procedimiento anterior, el equipo de guardaparques cuenta con un programa de monitoreo de especies en riesgo y de especies indicadoras de las condiciones del hábitat. Sin embargo, esta actividad no se utiliza como un medio para evaluar si las acciones de conservación han tenido un efecto o no. Una situación similar ocurre con los esporádicos estudios de la tasa de transformación de las cubiertas forestales hechas por especialistas (se han hecho dos en los últimos 20 años). La información generada, si bien permite hacer un diagnóstico de la situación en la que se encuentra la RB, no permite atribuir los cambios al éxito o fracaso de los proyectos implementados.

La OSC, por su parte, ha interactuado con evaluaciones de impacto de dos maneras. Por un lado, siguiendo lo convenido con los donantes en la planeación del proyecto, la OSC ha medido los indicadores propuestos en modelos lógicos, contando idealmente con una línea de base para establecer la magnitud del cambio (Cuadro 5). En ese escenario, las evaluaciones externas se han centrado en verificar: a) que los indicadores puedan medir impactos, b) que la organización tiene en marcha un sistema de evaluación que permite medirlos, y c) la veracidad de los datos reportados por la organización.

Cuadro 5. Indicadores de impacto extraídos de evaluaciones externas a los proyectos de la OSC.

Indicadores sociales	Indicadores ambientales
Porcentaje de incremento del ingreso de los beneficiarios.	Estimación de captura de carbono en reforestaciones y en el suelo.
Beneficios comunitarios generados a partir de los nuevos ingresos.	Superficie destinada exclusivamente a la conservación ⁷³
Cambios en actitudes hacia la conservación.	Reducción del índice de fragmentación de bosques.
Disminución de ilícitos ambientales.	Disminución de erosión de suelos.

⁷³ Al ser una RB, más allá de las zonas núcleo los propietarios tienen permitido realizar actividades agropecuarias. Por ello, un programa estratégico de la organización ha sido la compra de tierras y la oferta de compensaciones económicas para garantizar la conservación.

Percepción de mejoras en la calidad de vida a partir de incremento de servicios ambientales.	Dinámica poblacional de especies en riesgo.
--	---

Fuente: Elaboración propia

Dada su incursión en los últimos años en mecanismos de compensación por servicios ambientales, el sistema de evaluación de la OSC se ha enfocado en el monitoreo de la captura de carbono en bosques, específicamente encinares. Para ello, desarrollaron un protocolo local con la asesoría del servicio forestal de los Estados Unidos. Su propósito es incrementar la credibilidad de las compensaciones ofrecidas en el mercado voluntario de carbono.

De forma adicional, la OSC utiliza el retorno social de la inversión⁷⁴ (SROI por sus siglas en inglés) como metodología de evaluación de impacto (Cupitt, 2009). La organización ha utilizado la herramienta como parte del sistema de evaluación de algunos proyectos financiados por el BID y para estimar el efecto global que ha generado en la RB. Para ello, contrasta los ingresos generados para la economía local⁷⁵ y el valor anual de los servicios ambientales atribuibles a sus proyectos (Cuadro 6) con el recurso recibido a través de la gestión de financiamiento. Esto le ha permitido demostrar en diferentes evaluaciones un retorno de la inversión de entre 12 y 18 dólares por cada dólar invertido. La OSC actualiza la información a través de la aplicación de encuestas y entrevistas, reportes de visitas de seguimiento y trabajo de gabinete.

Cuadro 6. Algunos indicadores ambientales monetizados a través de la herramienta.

Indicador ambiental	Indicador financiero
Almacenamiento de carbono en la vegetación de la reserva	Valor en dólares por tonelada de carbono.
Servicios hidrológicos	Monto por hectárea que paga el gobierno mexicano en su programa de pago por servicios ambientales.
Servicios por biodiversidad	Monto por hectárea que paga el gobierno mexicano en su programa de pago por servicios ambientales.

Fuente: Elaboración propia a partir de documentos institucionales de la OSC

⁷⁴ La metodología, desarrollada a partir de un análisis tradicional de costo-beneficio, busca medir y comunicar el valor generado por una iniciativa mediante indicadores *proxy* para monetizar impactos sociales y ambientales. De esta manera, se calcula un coeficiente que relaciona el dinero invertido con el beneficio creado. El enfoque también integra indicadores cuantitativos, cualitativos y narrativos para expresar todas las dimensiones del cambio.

⁷⁵ Calculados a partir de a) los ingresos de proyectos productivos y turísticos apoyados por la OSC, b) las compensaciones económicas gestionadas para los propietarios de predios forestales y c) los salarios de su personal.

Su objetivo, en palabras del responsable de las evaluaciones, es comunicar el valor de una reserva de la biósfera en términos de su costo/beneficio. La intención es que la herramienta provea un soporte analítico para decisiones de financiamiento de políticos, inversionistas y organizaciones internacionales.

Factores ligados al uso de evaluaciones como herramienta de aprendizaje

En este apartado se analiza la capacidad de las organizaciones para utilizar las evaluaciones como una herramienta de aprendizaje (Tabla 7). En específico, se valora la influencia de criterios ligados a la calidad de las evaluaciones, variables organizacionales, y los arreglos institucionales que enmarcan la generación de evaluaciones.

Cuadro 7. Factores que promueven o inhiben el uso de evaluaciones como una herramienta de aprendizaje organizacional

Factores que promueven el uso de evaluaciones para aprender	Dependencia federal	Administración municipal	Universidad pública	OSC
Uso de retroalimentaciones provenientes de evaluaciones	X			X
Personal especializado en evaluaciones				X
Uso de modelos lógicos para establecer y monitorear indicadores de impacto.				X
Existencia de sistemas de gestión de la información	X			X
Factores que inhiben el uso de evaluaciones para aprender	Dependencia federal	AM	Universidad pública	OSC
Temor a las consecuencias de las evaluaciones	X	X	X	X
Inexistencia de modelos lógicos para establecer y monitorear indicadores de impacto.	X	X	X	
Falta de recursos (tiempo y dinero) para dar seguimiento a los proyectos	X	X	X	X

Rotación de personal		x		
----------------------	--	---	--	--

Fuente: Elaboración propia

Calidad de las evaluaciones

Un elemento básico de la calidad de una evaluación es su adecuación metodológica para generar información que conduzca a la toma de decisiones. Si el propósito es el aprendizaje organizacional, se necesita utilizar evidencia sobre la generación de cambios para adaptar los proyectos.

Cuando se desea conocer las razones por las que se consiguieron, o no, los cambios esperados, se ha resaltado la importancia de comenzar con la creación de modelos lógicos que representen las relaciones de causalidad esperadas entre las variables. Este enfoque facilita la identificación de los indicadores y la metodología con los que se evaluarán los resultados (Margoluis *et al.*, 2013; Pressey *et al.*, 2015).

Este aspecto ligado a la calidad de las evaluaciones es una falla esencial del sistema de evaluación de la AM, la universidad y la dirección de ANP. En documentos institucionales la AM afirma que sigue los criterios técnicos y administrativos de la Gestión basada en Resultados (GbR)⁷⁶, sin embargo esto no es así. Por un lado, en su plan de desarrollo sólo se planteaban objetivos generales⁷⁷, y por el otro los POA de las áreas no se utilizaban como referencia para evaluar el desempeño.

Por su parte, el proyecto del centro de investigaciones de la universidad no contaba con una teoría del cambio clara que permitiera definir sus alcances. A la fecha, sigue sin contar con un instrumento de planeación a partir del cual medir la obtención de resultados. Su plan estratégico sólo contempla indicadores de gestión (p. ej. número de publicaciones, capacitaciones, convenios y diagnósticos).

Además, en el caso de investigaciones aplicadas, como las que se realizan en el centro, no es posible equiparar la producción de conocimiento con su impacto. Para lograr los cambios en las condiciones sobre las que se pretende incidir, es necesario que el conocimiento generado tenga un efecto sobre el comportamiento de otros actores (NWO, s.f.). En general, es un error asumir que los grupos involucrados accederán, leerán, entenderán y utilizarán los artículos científicos para tomar decisiones (Salafsky *et al.*, 2001). Por ello, los productos académicos utilizados dentro del sistema de evaluación de la universidad no permiten establecer los cambios que ha conseguido el equipo de investigadores.

Con respecto del equipo de guardaparques de la CONANP, si bien un procedimiento organizacional solicita que el POA se elabore siguiendo la metodología del marco lógico, esto entra en contradicción con el catálogo de actividades y unidades de medida que se envía a los guardaparques para estandarizar la información reportada. Las métricas utilizadas miden avances en la implementación al aportar información descriptiva que documenta la

⁷⁶ Este enfoque establece la necesidad de orientarse hacia la generación de valor público, y no simplemente hacia el cumplimiento de procedimientos (González y Cerdán, 2008).

⁷⁷ P. ej. brindar apoyo con asesoría técnica a pequeños productores, promover el ecoturismo, o cuidar el medio ambiente mediante programas de reciclaje.

generación de productos comprometidos (p. ej. talleres de capacitación o número de comités operando). El programa de manejo del área tampoco ofrece orientación en ese sentido, pues la mayor parte de sus indicadores son de gestión.

En cambio, en el diseño de algunos de sus proyectos, la OSC sí utiliza teorías del cambio que distinguen entre actividades, resultados inmediatos y los cambios esperados. A partir de ellos, elabora indicadores, establece líneas de base, concreta una propuesta metodológica y solicita presupuesto para llevar a cabo el monitoreo y evaluación del proyecto. Lo anterior la coloca en una mejor posición para adaptar sus proyectos a partir de evaluaciones.

Es posible relacionar el desarrollo de sus capacidades para hacer evaluaciones con la gestión de proyectos multianuales financiados por donantes internacionales. Estos no solo le exigieron estándares de calidad en el diseño de proyectos, sino que le brindaron retroalimentaciones para perfeccionar su sistema de evaluación.

Variables organizacionales

Las organizaciones precisan tener claros los motivos y brindar los medios para que las evaluaciones ocurran. Entre estos se encuentran la disponibilidad de recursos, la contratación de personal o el desarrollo de procedimientos (Serrat, 2010).

Un conjunto de limitaciones identificadas por los integrantes de todas las organizaciones se relacionan con su capacidad para dar seguimiento a los proyectos emprendidos. En buena medida, la operación se centra en la ejecución de actividades, sin que se consideren acciones posteriores para verificar cambios y plantear adecuaciones. En general, esta es la lógica imperante en proyectos con una duración menor a un año, la cual afecta especialmente a las organizaciones gubernamentales, atadas a ejercicios fiscales. Además, los recursos con los que cuentan son escasos, por lo que perciben que destinar una parte a labores de monitoreo no es una inversión adecuada, pues es más valioso usarlo directamente en la ejecución de proyectos. Esta noción de sacrificio de actividades más urgentes para poder hacer evaluaciones, ya ha sido resaltada (Salzer y Salafsky, 2006).

Por otro lado, personal de la dependencia gubernamental y de la universidad señaló que la falta de tiempo es una condicionante que vuelve inviable el seguimiento. Los investigadores precisan impartir al menos quince horas de clase por semestre para poder concursar por una plaza de tiempo completo⁷⁸. Además, participan en diferentes tareas administrativas, como consejos académicos, tutorías o en la coordinación de diferentes áreas en los programas educativos. Esto deja poco tiempo para los proyectos de investigación y para otras actividades de vinculación que realizan (p. ej. capacitaciones o asesorías). Por su parte, los guardaparques realizan múltiples funciones debido a la falta de personal, lo cual dificulta el seguimiento de los proyectos. En parte esto se debe a que, a diferencia de otras dependencias que se apoyan en la contratación de profesionistas externos para asesorar las acciones de los programas de subsidio, con frecuencia los guardaparques absorben la carga de perfilar, acompañar y supervisar los proyectos.

⁷⁸ Los diez investigadores tienen contrataciones de 40 horas, pero actualmente solo dos de ellos tiene una plaza de tiempo completo.

Las únicas organizaciones que cuentan con puestos en la escala local especialmente dedicados a las evaluaciones son la AM y la OSC. A pesar de ser su atribución, la contraloría del municipio se enfoca en auditorías presupuestales. En cambio, la OSC se apoya en un especialista que asesora el diseño y operación de su sistema de evaluación. Su experiencia se concentra en el desarrollo de modelos lógicos, en la valoración económica de ecosistemas, y en el cálculo del retorno social de la inversión. De nuevo, la implementación de proyectos multianuales fue un factor determinante, al permitir la contratación de responsables de evaluación.

Además, esto abrió la necesidad en la OSC de contar con sistemas para gestionar información⁷⁹ que permitiera reportar avances en indicadores específicos. En contraste, el municipio y la universidad no cuentan con procedimientos para almacenar la información generada a través de la implementación de sus proyectos. Mientras tanto, la dirección de ANP comenzó a llevar un registro recientemente, al percatarse de que no contaba con información para evaluar el programa de manejo de la reserva. En este sentido, la dirección de ANP es la organización que más utiliza evaluaciones para identificar áreas de mejora en el desempeño de su personal y en el rediseño de sus instrumentos de planeación de largo plazo.

Arreglos institucionales

En última instancia, para explicar por qué las organizaciones desarrollan o no procedimientos vinculados a la generación de evaluaciones, se necesita comprender la influencia de arreglos institucionales sobre su comportamiento.

Tanto en el campo del desarrollo, como en el de la administración pública, el enfoque de la GbR es el paradigma que ha normado el desarrollo de reglas, procedimientos e incentivos vinculados con evaluaciones (Eyben, 2015). En el discurso, sus principios se encuentran presentes en las evaluaciones solicitadas por donantes y por dependencias de gobierno a la OSC, y también en la política federal que enmarca el sistema de evaluación de la CONANP, la AM, y la Universidad.

En el caso de la AM, su autonomía institucional le ha permitido mantener un déficit de rendición de cuentas, pues las evaluaciones ni siquiera existen como procedimiento organizacional. A pesar de que esta carencia quedó de manifiesto a través de la evaluación de la agenda para el desarrollo municipal, no se impulsaron acciones para modificar esta situación. Posiblemente esto se haya debido a un desinterés para implementar los cambios dado que restaba un año para el cambio de administración. Se ha señalado que el desarrollo de capacidades en evaluación requiere de un compromiso de largo plazo (Lennie *et al.* 2015). En definitiva, la constante rotación de personal en la AM es un desafío a considerar.

Más allá de este caso, resalta la discrepancia entre los principios de la GbR y la puesta en práctica de las evaluaciones. Con excepción de los proyectos multianuales desarrollados por la OSC, las evaluaciones se concentraron en la medición de resultados inmediatos. Esta situación limita un segundo supuesto de la GbR: que el conocimiento sobre el cambio generado permitirá orientar la práctica futura.

⁷⁹ Como sistemas de información geográfica y diferentes plantillas en hojas de cálculo donde se lleva el control de la gestión de los proyectos.

En la literatura sobre el tema, se ha dicho que la revisión externa de los proyectos para juzgar el desempeño merma el potencial de la evaluación como herramienta de aprendizaje organizacional (Ebrahim, 2005). Esto es así, pues el temor a consecuencias negativas en caso de no conseguir los objetivos erosiona la posibilidad de reflexionar en torno de reportes honestos que expongan resultados negativos (Mosse, 2004; Guijt, 2008). En la dirección de ANP, los guardaparques mencionaron que el temor a las consecuencias negativas de la evaluación de su desempeño ha generado el desplazamiento de objetivos. De esa manera, en ocasiones, han procurado el cumplimiento de metas o han propuesto actividades seguras aunque eso no implique la obtención de resultados de largo plazo.

Mientras tanto, coincidiendo con lo señalado para varias universidades del país (Gil, 2014; Gil y Contreras, 2017), la universidad ha transferido el enfoque evaluativo de la política científica federal a sus procedimientos de contratación y promoción laboral. Sin embargo, utilizar la producción de conocimiento científico como métrica de evaluación corre el riesgo de promover un comportamiento administrativo estratégico en los académicos, que los lleve a destinar más tiempo en cumplir con requisitos para tener estabilidad laboral o generar ingresos económicos adicionales, que en la ejecución de proyectos de largo plazo (Galaz y Gil, 2013; Alonso, 2014).

En contraste, las evaluaciones de proyectos multianuales financiados por agencias internacionales trascendieron las limitaciones anteriores. Por un lado, al solicitar la medición de indicadores de impacto y, por el otro, al incorporar evaluaciones periódicas para identificar aspectos susceptibles de mejora en la implementación de los proyectos. Se ha dicho que, impulsada a través de esquemas de rendición de cuentas, la agenda de la GbR puede contribuir a la profesionalización de las organizaciones al fortalecer o desarrollar sus sistemas de monitoreo y evaluación (Cousins *et al.*, 2015). En este caso, es evidente que la relación establecida entre la OSC y estas agencias fue un mecanismo a través del cual la organización institucionalizó su práctica evaluativa.

Dado el costo de llevar a cabo evaluaciones de impacto rigurosas que prueben la generación de beneficios sociales y ambientales, es probable que los donantes privilegien mecanismos indirectos para confiar en la obtención de resultados. Por ejemplo, análisis que garanticen que los proyectos están bien diseñados o la existencia de sistemas de evaluación en las organizaciones receptoras. Esta es la lógica que está detrás de evaluaciones *ex ante*, como la exigida por los estándares de Clima, Comunidad Y Biodiversidad la OSC, o de las diferentes metodologías para evaluar la efectividad en el manejo que utiliza la CONANP para cumplir con la meta 11 de Aichi. Sin embargo, diferentes investigaciones han señalado que una buena planeación y manejo de proyectos, si bien son condiciones necesarias, no son suficientes para asegurar la efectividad. Por ello, no deberían de sustituir a las evaluaciones de impacto (Carranza *et al.*, 2014; Pressey *et al.*, 2017).

Además, es importante considerar que las evaluaciones podrían ser adoptadas por las organizaciones para manejar la forma en la que son percibidas por audiencias externas, y no como un medio para tomar decisiones basadas en evidencia (Mayne, 2006; Okail, 2009). Esto podría deberse más a deficiencias en la forma en la que funciona el campo del desarrollo, donde existe una presión por conseguir forzosamente los objetivos, a pesar de la incertidumbre existente en la implementación de los proyectos (Guijt, 2008). Por ello, siguiendo estas “reglas del juego”, las organizaciones entienden la importancia que tienen las

evaluaciones para comunicar historias de éxito, pues de ello depende el flujo de recursos para implementar su agenda (Mosse, 2004).

Esta situación limita esquemas de evaluación orientados hacia el aprendizaje, donde es tan importante explicar el porqué de las fallas como el de los logros. Por ejemplo, la OSC asegura en diferentes documentos institucionales que su análisis SROI le brinda información para tomar decisiones con un enfoque adaptativo. Sin embargo, el énfasis puesto en la valoración del impacto a través de indicadores monetizables, visible en las últimas actualizaciones de su retorno social, permite observar que su propósito se centra en demostrar resultados para facilitar su gestión de recursos, no en el aprendizaje.

Cabe preguntarse si la sobresaliente capacidad de gestión financiera de la OSC se vería afectada en caso de que sustituyera una narrativa de éxito, por una historia matizada sobre el proceso de cambio, en la cual incorpore fallas y desafíos. Sin duda, la sistematización de ese análisis sería de mucha utilidad, no solo para mejorar el diseño de los proyectos de la OSC, sino también para otras organizaciones que enfrentan retos similares en la reserva. Por su parte, en el caso de la CONANP, es importante que las evaluaciones de la efectividad en el manejo sirvan para que los guardaparques hagan cambios procedimentales, y no sólo para que la dependencia compruebe el cumplimiento de la meta 11 de Aichi. Dado que es un proceso en marcha en la dirección de ANP, valdría la pena estudiar sus efectos sobre el aprendizaje en el futuro.

Conclusiones

El manejo adaptativo de proyectos requiere evaluaciones que midan información valiosa⁸⁰ para que el personal operativo reflexione y mejore su práctica. En las evaluaciones formativas, el valor se encuentra en que se detecten procesos de cambio para corregir y compartir errores de la organización en los supuestos sobre lo que sucede (Argyris y Schön 1978: 29). Eso implica trascender la miopía de evaluaciones de desempeño que valoran resultados inmediatos que son fáciles de medir.

En ese sentido, la OSC es la organización que tiene el sistema de evaluación más robusto para establecer relaciones de causalidad, dado el uso recurrente de modelos lógicos, líneas de base e indicadores de impacto. Los requerimientos de proyectos multianuales financiados por agencias internacionales han impulsado el desarrollo de sus capacidades en evaluación. En primer lugar, al requerir el cumplimiento de estándares de calidad en el diseño de los proyectos. En segundo lugar, al promover la contratación de especialistas en evaluación, los cuales contaban con un presupuesto para coordinar actividades de monitoreo y evaluación. En tercer lugar, al interactuar con evaluadores externos para identificar áreas de mejora en la implementación de sus proyectos.

Su análisis SROI tiene el propósito de brindar evidencia sobre la obtención de resultados, con la finalidad de fortalecer sus estrategias de gestión financiera. Sin embargo, dado el

⁸⁰ El título del artículo se inspiró en un aforismo sobre la evaluación de la educación superior acuñado por Robert Birnbaum, el cual advierte sobre el riesgo de valorar lo que es medible cuando no se puede medir lo que es valioso (Birnbaum, 2001).

reconocimiento internacional con el que ya cuenta, se considera que la OSC tendría un margen de maniobra para incorporar con seguridad enfoques de evaluación orientados al aprendizaje, como reportes donde documente fallas y desafíos por superar (EWB, 2017). Ese conocimiento podría estimular conversaciones con otros actores para mejorar la práctica de la conservación y el desarrollo, convirtiéndose así en un mecanismo más a través del cual la OSC genera beneficios sociales.

En contraste, la AM carece de prácticas gerenciales que fundamenten la gestión de los proyectos sobre decisiones racionales. No hay forma de evaluar el desempeño ni la obtención de resultados, pues ni su plan de desarrollo ni los POA de las áreas contemplan indicadores para ello. Su rendición de cuentas se reduce a procedimientos para verificar un ejercicio adecuado del presupuesto y a informes donde se sintetizan las actividades realizadas.

Si bien, de manera voluntaria la AM se involucró en una evaluación para identificar áreas de mejora, producto de la cual se identificó la necesidad de incorporar indicadores en sus documentos de planeación, las recomendaciones no fueron implementadas. Dada la autonomía institucional del municipio, para que una cultura evaluativa se institucionalice es clave que las autoridades valoren las evaluaciones como una herramienta para alcanzar objetivos, y no como un mero ejercicio burocrático. Se considera que la demanda ciudadana de esos estándares organizacionales puede ser una clave para incentivar el proceso. En definitiva, un desafío será la rotación de personal y de liderazgo, pues interrumpe el desarrollo de capacidades en evaluación.

En un sentido similar, el centro de investigaciones requiere mejorar sus procedimientos evaluativos. Si la universidad desea asumir un rol de promotora de desarrollo regional, entonces es necesario reestructurar el sistema de evaluación del desempeño de sus investigadores. Su autonomía institucional le permitiría modificar el sistema de incentivos que dicta la política federal.

Por ejemplo, siguiendo la metodología de rutas de generación de impacto (Alvarez *et al.*, 2010), podría solicitar a los investigadores el desarrollo de modelos lógicos donde se describan las contribuciones sociales que se busca generar con la investigación, los mecanismos de interacción con otros actores y los supuestos bajo los cuales el modelo funciona. Concebir las investigaciones de esta manera permite incorporar la evaluación como una herramienta de mejora continua, puesto que los resultados obtenidos permitirían refinar la teoría del cambio original.

La CONANP cuenta con el sistema de evaluación más diversificado, pues cuenta con procedimientos que le permiten evaluar el desempeño, identificar áreas de mejora organizacional y valorar la obtención de resultados de conservación. Sin embargo, a nivel local la dirección de ANP no cuenta con indicadores para medir cambios, ni en su POA, ni en el programa de manejo de la reserva. La metodología i-efectividad tiene el potencial para promover mejoras en el manejo que los guardaparques hacen. A futuro una forma de verificar su uso será la integración del monitoreo biológico como mecanismo de evaluación de sus proyectos.

Es muy probable que las restricciones presupuestales de la dependencia continúen, así como los requerimientos de reportar tan sólo el cumplimiento de resultados inmediatos. En ese contexto, tendrá un papel fundamental el liderazgo del director de la reserva, el cual podría tejer alianzas con el resto de las organizaciones en torno del nuevo programa de manejo de

la RB, actualmente en proceso de actualización. El documento rector de planeación estratégica facilitaría la integración de un sistema de medición compartido, el cual sirva para retroalimentar los avances hacia metas comunes de conservación y desarrollo.

Capítulo siete

¿Qué necesita cambiar para permitir el aprendizaje a partir de evaluaciones?

Esperanza no es lo mismo que optimismo. No es la convicción de que algo saldrá bien, sino la certeza de que algo tiene sentido, independientemente de cómo resulte.

-Václav Havel-

En este apartado se sintetizan las principales contribuciones de la tesis, entremezcladas con vetas de exploración futura y recomendaciones específicas para fortalecer la capacidad evaluativa de las organizaciones estudiadas. El objetivo de la investigación fue comparar la forma en que distintas organizaciones hacen y usan evaluaciones de su propio trabajo en una reserva de la biosfera. Esto con la intención de establecer similitudes y diferencias en el grado que utilizan evaluaciones como herramienta de aprendizaje organizacional. El fenómeno se describió en cuatro organizaciones que gestionan proyectos vinculados con el manejo de recursos naturales: una administración municipal, una dependencia del gobierno federal, una universidad pública y una organización de la sociedad civil.

A continuación, se estructuran las respuestas dadas a cada una de las preguntas de investigación. Estas, a su vez, se relacionan con las brechas en el conocimiento sobre el tema y los objetivos específicos. En primer lugar, se presentan las aportaciones ligadas a los dos casos centrales de la investigación, y después, la información relativa a la comparación de los cuatro casos.

7.1 CASO DE ESTUDIO 1- CONANP

Brecha de investigación

No se ha estudiado el sistema de evaluación diseñado por la CONANP. A su vez, se desconoce la influencia que tiene sobre las decisiones de los equipos de guardaparques que administran las áreas protegidas. Esta escala de análisis es relevante, pues los guardaparques deberían ser los usuarios principales de evaluaciones que les permitan manejar sus proyectos a través de adaptaciones.

Objetivo específico

Identificar los factores que promueven e inhiben el uso de evaluaciones para dirigir la toma de decisiones de los guardaparques en una reserva de la biosfera del centro de México.

Preguntas de investigación

En la dependencia federal ¿Cómo se hacen y usan evaluaciones relacionadas con el manejo del ANP?

¿En qué grado las evaluaciones se han utilizado para el manejo adaptativo del ANP?

¿Cuáles son las principales barreras para que los guardaparques aprendan a partir de esas evaluaciones?

La investigación analizó el sistema de evaluación de la CONANP a la luz de su uso por los guardaparques de la reserva. Se identificó que estos interactúan con tres tipos de evaluación. En primer lugar, una evaluación del desempeño que verifica el cumplimiento de la planeación operativa anual que los propios guardaparques se trazan al inicio del año. En segundo lugar, una evaluación periódica de la efectividad del programa de manejo del área, documento de planeación estratégica multianual.

Finalmente, un par de evaluaciones de la efectividad en el manejo que buscan identificar brechas en los recursos y procedimientos que emplea el equipo de guardaparques.

En términos generales, la evaluación del desempeño es utilizada como un mecanismo de rendición de cuentas intraorganizacional que permite a la dependencia: a) agregar información de los resultados inmediatos obtenidos por las distintas direcciones de ANP, b) estandarizar algunas actividades a realizarse en las áreas y c) monitorear el cumplimiento de las metas (*outputs*) que los propios guardaparques seleccionaron con base en su experiencia y las directrices del programa de manejo del área.

Mientras tanto, en términos formativos, el procedimiento estimula en los guardaparques un ejercicio reflexivo que les permite expresar el aprendizaje informal que han obtenido, utilizándolo como un insumo para la elaboración del programa operativo del siguiente año. Debido a que la evaluación del POA no involucra la medición del efecto de las acciones, este aprendizaje no se basa en evidencias sobre la obtención de resultados. En cambio, se basa en las experiencias no sistematizadas de los guardaparques, insumo con el cual deciden hacer ajustes en sus planeaciones.

Por su parte, la cédula de evaluación de la efectividad del programa de manejo, más allá de un ejercicio de rendición de cuentas, tiene el objetivo explícito de justificar la pertinencia de actualizar el documento de planeación estratégica del área. Dada la antigüedad del programa de manejo (más de 15 años), la decisión de actualizarlo ya había sido tomada por los guardaparques, por ello, vertieron en el documento razonamientos basados en su experiencia. Independientemente, experimentar el proceso de la evaluación les permitió identificar áreas de mejora en sus procedimientos de registro, pues tuvieron dificultades para rastrear información sobre las acciones realizadas a lo largo de los años.

Con respecto de las evaluaciones formativas, el equipo de guardaparques interactuó recientemente con un par de evaluaciones de la efectividad en el manejo, cuyo énfasis, en teoría, es identificar áreas de mejora procedimental que, al integrarse, impacten positivamente en los resultados que generan. La primera, INDIMAPA, fue considerada por los guardaparques como parte de un diagnóstico global de la situación de las direcciones de área en el país, por lo que la información no ocasionó cambios ni en su forma de trabajo ni en los recursos disponibles. La segunda, i-efectividad, fue desarrollada ex profeso por la dependencia para retroalimentar el accionar de los guardaparques. Sin embargo, por ser una metodología de reciente implementación, en esta investigación no pudo estudiarse su efecto como herramienta de aprendizaje. Por ello, documentar la interacción de los guardaparques con esa metodología es una interesante área de exploración futura.

Para ello, habría que identificar si el uso periódico de esta metodología *ad hoc* de evaluación causa cambios en los procedimientos empleados así como en el tipo de acciones registradas en los POA. De cualquier forma, no se podrá medir el efecto de estas adecuaciones sobre la efectividad, a menos de que el equipo de

guardaparques incorpore indicadores de resultado (*outcomes*) en sus planeaciones. Actualmente, ni siquiera el programa de manejo plantea indicadores que vinculen las actividades con efectos concretos a generar. Por ese mismo motivo, el monitoreo de especies clave y el análisis periódico de cambios de uso de suelo, si bien brindan información diagnóstica, no pueden emplearse para establecer el valor público generado por la dirección de ANP.

En este sentido, la falta de teorías del cambio vinculables a sus planes de trabajo es una de las barreras para que los guardaparques realicen un manejo adaptativo del área. Por este motivo, sería conveniente que los guardaparques establezcan en su POA los cambios más próximos que esperan observar como resultado de cada una de sus acciones. Actualmente, los guardaparques describen *outputs* (p. ej. # talleres realizados, # reuniones realizadas) y señalan el momento del año en que generarían la evidencia, pero no indican cómo notarían que, por ejemplo, las personas involucradas desarrollaron habilidades o que avanzaron en el establecimiento de marcos regulatorios. La intención es que puedan hacer una medición de bajo costo de su progreso al enfocarse en resultados cercanos a la implementación, como cambios de comportamiento esperados en los actores con los que interactúan. Este monitoreo recurrente proveería oportunidades para aprender de los errores y hacer ajustes. Además, desarrollar la teoría del cambio de sus estrategias sería un excelente ejercicio pedagógico en sí mismo, pues esta les ayudaría a clarificar supuestos y a orientarse hacia el cumplimiento de objetivos de mediano plazo.

Una fortaleza presente en varias de las evaluaciones de la CONANP, es que asumen que los guardaparques serán los productores y usuarios de la información generada. Esta concepción de las evaluaciones, fincadas en los intereses y la posición epistémica preferencial del personal operativo, es una buena base para manejar los proyectos a través de adaptaciones.

Aun así, aunque las evaluaciones sean un estímulo que fomenta la reflexión, es necesario que el aprendizaje se sistematice, transparentando las razones que fundamentan los cambios a realizar. Actualmente, por ejemplo, los guardaparques solo reportan el POA finalizado y si bien las lecciones aprendidas forman parte del proceso, éstas no quedan documentadas. Por estos motivos, sería conveniente integrar en la metodología algunas técnicas de gestión del conocimiento que permitan visibilizar el aprendizaje acumulado. Por ejemplo, serían útiles el reporte de fallas (EWB, 2017), notas de progreso mensual (Serrat, 2010), o la indagación apreciativa (Catsambas, 2016).

A escala organizacional, la incapacidad de los guardaparques para dar seguimiento a los cambios generados por las acciones que realizan es otra de las barreras para manejar el ANP a través de adaptaciones. Lo anterior se debe tanto a la falta de recursos destinados para esa tarea, como a la saturación de actividades a realizar por los guardaparques.

En un contexto de carestía presupuestal como el que desde hace años enfrenta la dependencia, es evidente que la balanza entre recursos para operar y recursos para evaluar se incline por la primera actividad. Por ello, y dado que no se avizora un

cambio en esta situación, adquiere una relevancia mayor la integración de evaluaciones formativas que utilicen de forma sistemática la experiencia de los guardaparques para hacer sentido del contexto y adaptar sus decisiones.

Independientemente, sería recomendable que los guardaparques establezcan alianzas con otras organizaciones para poder medir indicadores de impacto vinculables a sus acciones. Una fuente importante de información podría provenir de organizaciones académicas, las cuales utilicen su capital humano⁸¹ para generar investigaciones orientadas a decisiones.

Más aún, dado que en la reserva confluyen múltiples actores, sería deseable que los guardaparques tejan alianzas en torno del cumplimiento de los objetivos del programa de manejo del área. Para ese fin, una de las recomendaciones derivadas del foro realizado (Anexo 3) fue fortalecer la figura del consejo asesor, el cual es un órgano de participación ciudadana promovido por la dependencia en áreas naturales protegidas. Como ese, en las áreas rurales de México existen varios espacios deliberativos que promueven el intercambio de opiniones entre actores sociales e institucionales. En general, su propósito es generar consensos y mejorar la calidad de las decisiones públicas. En ese sentido, una veta interesante de estudio futuro está en el análisis de su funcionamiento en la reserva: ¿Qué tipo de decisiones se toman?, ¿cuáles son los resultados que se espera obtener?, ¿cómo se sistematiza y da seguimiento a los acuerdos?, ¿pueden establecerse compromisos para la acción conjunta? De lograr influir en su funcionamiento, el consejo podría ser el punto de partida de un sistema de evaluación compartido, el cual integre una visión común, facilite el aprendizaje conjunto y permita ahorrar costos en la medición de indicadores clave.

Finalmente, la tensión entre lógicas sumativas y formativas de evaluación es una barrera a trascender para que los guardaparques usen evaluaciones en el manejo adaptativo del área. Por un lado, es innegable el interés de la dependencia por integrar herramientas de evaluación orientadas al aprendizaje, evidencia de ello son el desarrollo de la metodología i-efectividad con la asesoría de Marc Hockings, o la integración de herramientas recomendadas por la Agencia Alemana para la Cooperación Internacional (GIZ), como la georreferenciación de actividades o la metodología de gestión Capacity Works. Una de las influencias institucionales que han promovido estas adopciones ha sido el cumplimiento de la Meta 11 de Aichi, mecanismo de rendición de cuentas que facilitaría la gestión de financiamiento externo para la dependencia.

Sin embargo, la lógica anterior coexiste con procedimientos que dificultan su adopción. Los guardaparques reportaron modificaciones en su comportamiento como resultado del temor a las evaluaciones cuantitativas de su desempeño, por ejemplo, al privilegiar resultados de corto plazo que fueran fáciles de conseguir. El riesgo es que este sesgo de contabilidad, presente con frecuencia en los enfoques de gestión basada en resultados de la administración pública, ocasione una perversión en el aprendizaje de los guardaparques, haciéndose mejores en la selección y cumplimiento de metas para ser mejor evaluados y no necesariamente

⁸¹ P. ej. especialistas en facilitación, percepción remota o monitoreo biológico.

para desempeñarse mejor en el mundo real.

Es importante tener esto en cuenta cuando se analicen los efectos de i-efectividad. Actualmente, como resultado final de la integración de múltiples indicadores, la metodología arroja un índice que pretende reflejar la efectividad en el manejo. Por ello, para prevenir su uso exclusivamente para rendir cuentas y no para el manejo adaptativo de las áreas, sería conveniente complementar ese número al invitar a que los guardaparques reporten fallas, el grado en el que tomaron decisiones basadas en lo aprendido, y cómo esto permitió responder mejor al contexto del área que manejan.

7.2. CASO DE ESTUDIO 2- UNIVERSIDAD PÚBLICA

Brecha de investigación

En México, no se ha estudiado la forma en que las universidades públicas evalúan sus proyectos de investigación aplicada. El énfasis se ha puesto, en cambio, en el sistema de evaluación de la política científica federal y su influencia sobre el comportamiento de los investigadores.

Objetivo específico

Identificar los factores vinculados a la producción y uso de evaluaciones en un centro público de investigación agrícola.

Preguntas de investigación

En la universidad ¿Cómo se hacen y usan evaluaciones en un centro público de investigación agrícola inserto en la reserva?

¿Qué factores influyen en el propósito de las evaluaciones utilizadas por la universidad?

¿Cómo influye el sistema de evaluación sobre las decisiones de los investigadores?

Dado que el aprendizaje y el conocimiento son valores fundamentales para las universidades, se podría asumir que las evaluaciones tienen un lugar privilegiado en este tipo de organización. Sin embargo, la investigación identificó que las evaluaciones asociadas a los proyectos del Centro Público de Investigación Agrícola (CPIA) no permiten estimar cambios ni retroalimentar los proyectos con la finalidad de mejorarlos. Esta limitación es reconocida por la propia organización, pues una revisión de su estrategia para orientar la investigación hacia la resolución de problemas sociales concluyó que la institución no cuenta con mecanismos para medir el impacto de los proyectos. Las evaluaciones identificadas tuvieron tres propósitos: 1) verificar la calidad de los proyectos antes de su aprobación, 2) rendir cuentas sobre los resultados inmediatos obtenidos con los proyectos y 3) tomar decisiones laborales con base en el desempeño del personal académico.

En torno del primero de los propósitos, sea como requisito para recibir el financiamiento por agencias públicas de investigación, o como un mecanismo interno de la universidad para registrar proyectos y priorizar la entrega de fondos para realizar proyectos, se identificó el uso de evaluaciones *ex ante* en las que pares académicos revisan que la propuesta cumpla con estándares de calidad científica. Su uso prioritario, por tanto, es la dictaminación de propuestas y no retroalimentar el diseño de los proyectos con miras a su mejora.

Desde esta lógica, una mejora procedimental que podría llevar a cabo la universidad es la siguiente. Actualmente, la universidad maneja un protocolo estandarizado que solicita distinguir si la investigación es básica o aplicada y, además, señalar de forma general los resultados esperados. En caso de que la investigación sea aplicada, como criterio de elegibilidad podría requerirse ir más allá al describir de forma detallada la ruta de generación de impacto a través de la que el conocimiento tendría consecuencias prácticas. Esto permitiría, a su vez, justificar la solicitud de presupuesto específico para actividades de intermediación de conocimientos.

Por motivos de transparencia, pero también de aprendizaje, sería conveniente entregar los dictámenes de las propuestas no seleccionadas. Además, si se deseara que la evaluación *ex ante* sirviera para retroalimentar los proyectos a financiar, con miras a mejorar su diseño y maximizar sus posibilidades de generar resultados, sería adecuado promover encuentros donde los revisores dialoguen con los investigadores que propusieron el proyecto seleccionado.

Un aspecto a destacar en relación a las evaluaciones *ex ante* que utiliza la universidad, es la existencia de comités de bioética que verifican el cumplimiento de procedimientos para garantizar la protección de los grupos que pudieran resultar perjudicados, ejemplo de esto es la solicitud del Consentimiento, Previo, Libre e Informado (CPLI). Al menos en teoría, si se sigue de forma correcta el proceso para la obtención del CPLI, entonces se promovería la participación de los actores involucrados desde el diseño mismo de los proyectos. Por ese motivo, un área interesante de exploración futura es cotejar su uso en la práctica. ¿De qué forma se obtiene el CPLI?, ¿El proceso influye en el diseño de los proyectos?, ¿El proceso concluye con la entrega de resultados a los grupos involucrados?

En torno del segundo de los propósitos, se identificó el uso de evaluaciones de desempeño que basaban la rendición de cuentas en la entrega de productos previamente comprometidos. Por un lado, la evaluación del proyecto con financiamiento externo que dio origen al CPIA verificó la obtención de una gama de resultados inmediatos que buscaban sentar las bases de la operación del Centro (p. ej. un plan estratégico, un calendario de capacitación continua, una estrategia para atender las demandas de los productores). Si bien, el uso de mecanismos de intermediación como *proxy* de la incidencia es aconsejable para evaluar proyectos en los que, por problemas de latencia, es difícil estimar el impacto en el corto plazo, hay que tener claro que su existencia no implica la obtención de resultados.

Para saberlo, es necesario que el equipo de investigadores del CPIA evalúe la efectividad de las estrategias de interacción planteadas en sus planes y manuales (¿las unidades demostrativas conducen a la adopción de las técnicas?, ¿la

adopción de las técnicas incrementaron la rentabilidad de la unidad de producción?), proceso de aprendizaje que les permitiría hacer ajustes en el diseño de los proyectos. Sin embargo, la investigación mostró una brecha en la implementación de evaluaciones orientadas al aprendizaje, las cuales se plantearon en el papel para retroalimentar el plan estratégico del CPIA, pero no se llevaron a cabo. A pesar de que el proceso fue impulsado por la Dirección de Planeación de la Universidad, prevaleció un modelo de gestión individual de los proyectos que no requería a los investigadores analizar de forma conjunta los resultados obtenidos.

Debido a estas inercias de trabajo disciplinar, y dado que el CPIA es una apuesta organizativa de la universidad para atender problemas complejos de forma interdisciplinaria, es importante identificar qué mecanismos pueden detonar la colaboración entre los investigadores. En este escenario, una interesante área de exploración futura sería analizar el grado en que los líderes de este tipo de Centros, respaldados por sistemas de soporte de su organización, pueden usar evaluaciones formativas recurrentes de planeaciones y proyectos como un medio para acercar al equipo hacia las estructuras interdisciplinarias deseadas.

Lejos de este escenario, el procedimiento de evaluación *ex post* mandatado por la universidad se basa en la entrega de productos académicos comprometidos al registrar el proyecto, sin hacer diferencias entre proyectos de investigación y vinculación. Sin embargo, estos indicadores no registran en sí mismos los efectos del conocimiento generado, lo cual, dada la orientación práctica de las investigaciones en la interfaz sociedad-ambiente, es especialmente importante para evaluar la efectividad del CPIA.

Si bien, es posible que los productos académicos contengan esa información, no puede garantizarse que es así. Con ese sistema de reporte, las consecuencias prácticas de las investigaciones se ocultan detrás de números. Además, la falta de otro tipo de evaluación invisibiliza el proceso de aprendizaje y adaptación que pueden experimentar los investigadores al implementar los proyectos. En el proceso, es posible que exista aprendizaje de doble ciclo, pero actualmente no es mediado por evaluaciones.

Por ese motivo, se recomienda que los investigadores del CPIA integren teorías del cambio en el diseño mismo de sus proyectos. Ese ejercicio analítico les permitiría identificar a los actores con los que van a interactuar, las estrategias de intermediación que emplearán, y los indicadores con los que identificarán los cambios. Esta representación causal sería el punto de partida de reuniones recurrentes de autoevaluación en las que los investigadores sistematicen los avances conseguidos, las fallas detectadas y los ajustes a realizar.

Para sostener estas recomendaciones en sistemas de soporte, las mejoras podrían anclarse al sistema de gestión de la calidad utilizado por la Dirección de Investigación de la Universidad. De esta manera, las teorías del cambio se requerirían como parte del proceso de registro y evaluación *ex ante*; mientras que los reportes de autoevaluación serían parte de evaluaciones intermedias y *ex post*. Además, podría aprovecharse el sistema de capacitación con el que la universidad ya cuenta, como plataforma para desarrollar las capacidades de evaluación de los

investigadores que realizan proyectos de ciencia aplicada. Esto no solo proveería las herramientas, facilitando su uso, sino que también promovería la valoración de los cambios al justificar su pertinencia.

Por último, en torno del tercero de los propósitos, la universidad utiliza un sistema cuantitativo para valorar el desempeño de los investigadores en los concursos para obtener plazas, estímulos y promociones laborales. Si bien en los tabuladores también se valoran otros criterios, en los últimos años se ha incrementado el peso de indicadores de capacidad y productividad académica que se encuentran presentes en los programas federales que evalúan y dotan de recursos extraordinarios a la universidad. Por ende, aunque goza de autonomía institucional, los procedimientos de contratación y promoción laboral de la universidad se encuentran muy influenciados por el enfoque evaluativo de la política científica federal.

Esta alineación tiene varias consecuencias. En primer lugar, existe una disociación entre la teoría de evaluación adoptada por la Universidad (Responsabilidad Social Universitaria) y el sistema de evaluación en uso, el cual se apoya en una concepción de la rendición de cuentas basada en la productividad académica, restringiendo la valoración a la comunidad científica encargada de asegurar la calidad de las investigaciones. Sin embargo, en el marco de la Responsabilidad Social Universitaria, modelo de gestión elegido por la organización, es necesario que la universidad rinda cuentas con base en los impactos generados por sus proyectos de investigación aplicada. Por ese motivo, la universidad requiere adoptar enfoques de evaluación que le permitan hacer juicios sobre el mérito, pero también sobre la relevancia de las investigaciones. La valoración del mérito se enfocaría, como ahora, en que los proyectos cumplan con estándares científicos, y el análisis de la relevancia pondría atención en el valor extrínseco de los proyectos, es decir, en la indicación de que el conocimiento tuvo consecuencias de algún tipo para atender problemas sociales.

En segundo lugar, el sistema cuantitativo de evaluación del desempeño envía el mensaje a los investigadores de que más es mejor, lo que puede incidir en un comportamiento orientado a publicar como meta en sí misma. Pudo identificarse evidencia de esto en una estrategia de colaboración expuesta a los investigadores del CPIA por un académico externo. Al proponerse la conformación de un Cuerpo Académico, la lógica dominante era impulsar la productividad, sin que fueran mencionados los procesos de planeación estratégica para definir los problemas regionales a atender de forma conjunta.

Los investigadores del CPIA reconocieron una tensión entre cumplir con indicadores académicos que les brinden mejoras laborales (estabilidad, ingresos) e involucrarse en procesos sociales y políticos que incrementen el impacto de sus proyectos. Si los investigadores invierten recursos en actividades de intermediación, pero no cambia el sistema para evaluar su desempeño, entonces incurrirían en costos que ralenticen el avance de su carrera académica. Por ello, es crucial transformar las reglas del juego al adaptar la cultura de evaluación científica de la universidad. En un primer nivel, la universidad podría incorporar mecanismos que incentiven los cambios, como esquemas de financiamiento focalizado para atender

desafíos tecnológicos, sociales y económicos específicos. Esos fondos podrían funcionar como un piloto para incorporar instrumentos de evaluación *ex ante* y *ex post* que promuevan que los investigadores, más allá del cumplimiento de los estándares metodológicos de su disciplina, formen equipos interdisciplinarios, inviertan en actividades de intermediación del conocimiento, y reflexionen sobre los cambios conseguidos.

En un segundo nivel, la universidad podría hacer cambios en su reglamento para la contratación y promoción del personal académico. Esto, sin duda, incentivaría el cumplimiento de estándares de desempeño deseados por la organización. Entre las prácticas recomendadas, se encuentran el complementar los criterios cuantitativos utilizados para tomar decisiones de obtención de plazas, de manera que, al valorar la trayectoria de un investigador, el tipo de revista o la cantidad de publicaciones sean sólo una variable dentro de una amplia gama de indicadores que reflejen la relevancia de su trabajo (Cagan, 2013). Sin embargo, si se usan valoraciones cualitativas, es importante dejar claro y demostrar de forma pública cómo se atienden posibles sesgos, discriminación y otras formas de trato injusto (Science Europe, 2020). Además, dado que no existe un modelo de evaluación que pueda aplicarse a todas las áreas de la ciencia, es necesario evaluar las investigaciones en su propio contexto. Para ello, habría que desarrollar indicadores *ad hoc* para cada disciplina científica (Hicks *et al.*, 2015). Con ese fin, la universidad podría promover encuentros por facultad en donde se identifiquen desafíos, áreas de oportunidad y mejores prácticas en la evaluación de proyectos propios de las disciplinas alojadas en esas unidades académicas. Considerando lo anterior, se generó una propuesta para la administración de rectoría 2021-2024 (Anexo 4).

7.3 COMPARACIÓN INTERORGANIZACIONAL DEL USO DE EVALUACIONES

Brechas de investigación

A nivel internacional, se han hecho llamados para comprender mejor cómo se manifiesta en la práctica el manejo adaptativo en el campo del desarrollo. A su vez, se ha solicitado incrementar la base de evidencia sobre las condiciones necesarias para institucionalizarlo, de forma que logre sus objetivos de iteración programática y toma de decisiones basadas en evidencia. Por ello, se requieren más estudios de caso que ilustren el fenómeno.

En México, el estudio sobre el uso de evaluaciones se ha concentrado en la política de evaluación del gobierno federal. Sin embargo, existen pocas investigaciones sobre el uso de evaluaciones en la escala local, con organizaciones directamente involucradas en la implementación de proyectos de conservación y desarrollo. Es importante estudiar el fenómeno a ese nivel, pues los integrantes que operan las intervenciones deberían ser los principales usuarios de evaluaciones para manejar proyectos a través de adaptaciones.

Objetivos específicos

Preguntas de investigación

Realizar un análisis comparativo de la forma en que distintos tipos de organización hacen y usan evaluaciones en una reserva de la biosfera del centro de México.

Establecer el grado en que las organizaciones han adoptado evaluaciones como una herramienta de aprendizaje organizacional.

¿Cómo se hacen y usan evaluaciones en cuatro organizaciones distintas con proyectos en la reserva?

¿Cuáles son sus semejanzas y diferencias?

¿Qué factores promueven o inhiben el uso de evaluaciones como una herramienta de aprendizaje organizacional?

La investigación sistematizó los diferentes tipos de evaluación con los que interactúan las organizaciones (Cuadro 29). Los resultados muestran diferencias importantes entre ellas. Mientras que la Dirección de ANP y la OSC tienen los sistemas de evaluación más diversificados, cubriendo propósitos ideales de rendición de cuentas (RC), aprendizaje organizacional (AO) y gestión de financiamiento (GF), la universidad pública emplea un sistema de evaluación más limitado en alcances y, en el caso de la administración municipal este es prácticamente inexistente. Por ello, a nivel de los artefactos per se, la dependencia federal y la OSC están mejor posicionadas para aprender a partir de evaluaciones.

Cuadro 29. Jerarquización de las organizaciones en función del potencial de sus evaluaciones como herramienta de aprendizaje.

	ORGANIZACIÓN	TIPO DE EVALUACIÓN	PROPÓSITO		
1	OSC	Evaluación <i>ex ante</i>	AO	GF	
		Evaluación de desempeño	AO	RC	
		Evaluación de impacto	AO	RC	GF
	Dirección de ANP	Evaluación de desempeño	AO	RC	
		Evaluación de la efectividad organizacional	AO	RC	GF
		Evaluación de impacto	AO		
2	Universidad Pública	Evaluación <i>ex ante</i>	Dictaminación de proyectos		
		Evaluación de desempeño	RC		
	Administración municipal	Evaluación de la efectividad organizacional	AO		

AO. Aprendizaje organizacional; RC. Rendición de cuentas; GF. Gestión de financiamiento

Fuente: Elaboración propia

En diversos documentos, las cuatro organizaciones hacen referencia a la importancia de adoptar una gestión basada en resultados como medio de rendición de cuentas. Para todas, las evaluaciones son el medio para materializarlo. Mientras tanto, en términos de aprendizaje, la OSC, la universidad pública y la CONANP refirieron en su discurso la importancia de manejar los proyectos a través de adaptaciones, situando de nuevo a las evaluaciones como un medio para conseguirlo.

La investigación identificó divergencias entre estas intenciones sobre el uso que podrían tener las evaluaciones y su uso real. En la práctica, existen una serie de factores que dan forma a los sistemas de evaluación de las organizaciones, los cuales, explican por qué existen desconexiones entre la teoría adoptada y la teoría en uso. Estos se sistematizaron distinguiendo tres niveles, el metodológico, el organizacional y el institucional (Cuadro 30).

Cuadro 30. Factores que promueven o inhiben el uso de evaluaciones como una herramienta de aprendizaje en las organizaciones estudiadas

ORGANIZACIÓN	Factores que habilitan el uso de evaluaciones formativas		
	Institucionales	Organizacionales	Metodológicos
OSC	Estándares de evaluación requeridos por donantes internacionales	Personal especializado en evaluación Sistemas de gestión de la información	Uso del proceso Uso de modelos lógicos para establecer y monitorear indicadores de impacto.
Dirección de ANP	Cumplimiento de la meta 11 de Aichi	Sistemas de gestión de la información	Uso del proceso
Universidad Pública	--	--	--
Administración Municipal	--	--	--
ORGANIZACIÓN	Factores que inhiben el uso de evaluaciones formativas		
	Institucionales	Organizacionales	Metodológicos
OSC	Sesgo de contabilidad Miedo a fallar	Falta de recursos (tiempo y dinero) para seguimiento	--
Dirección de ANP	Sesgo de contabilidad Comportamiento administrativo estratégico Miedo a fallar	Falta de recursos (tiempo y dinero) para seguimiento	Inexistencia de modelos lógicos para establecer y monitorear indicadores de impacto.
Universidad Pública	Sesgo de contabilidad Comportamiento administrativo estratégico	Falta de recursos (tiempo y dinero) para seguimiento	Inexistencia de modelos lógicos para establecer y monitorear indicadores de impacto.

	Miedo a fallar		
Administración Municipal	Déficit de rendición de cuentas Miedo a fallar	Rotación de personal Falta de recursos (tiempo y dinero) para seguimiento	Inexistencia de modelos lógicos para establecer y monitorear indicadores de impacto.

Fuente: Elaboración propia

7.3.1 Nivel metodológico

La OSC fue la única organización que empleó modelos lógicos en el diseño de proyectos, específicamente a través de la metodología del Marco Lógico y la Teoría del Cambio. A través de ellas, distinguió indicadores para identificar el cumplimiento de resultados inmediatos y cambios esperados, lo cual la situaba en una mejor posición para detectar procesos y adaptar sus proyectos.

En contraste, los sistemas de evaluación de la Dirección de ANP, la Universidad Pública y la Administración Municipal adolecen de este estándar ligado a la calidad de las evaluaciones. Aunque entre los procedimientos de la CONANP se establece que el POA de las direcciones de ANP debería hacerse con la metodología del Marco Lógico, en la práctica el formato estandarizado que se solicita a los guardaparques solo precisa que se definan los *outputs* cuantificables que resulten de las actividades. Este problema se extiende al Programa de Manejo del ANP, el cual tampoco provee una visión de largo plazo basada en lógicas causales.

Por su parte, como ya ha sido expuesto, la adición de este tipo de modelos lógicos fortalecería la orientación a resultados de las investigaciones aplicadas realizadas en universidades públicas, dado que actualmente la rendición de cuentas se concentra en la productividad académica. Finalmente, la administración municipal, organización con las menores capacidades demostrables de evaluación, tampoco establece en ningún documento la vinculación entre un problema público, los mecanismos para atenderlo y metas a conseguir medidas con indicadores de gestión (*outputs*) y resultado (*outcomes*). En cambio, para rendir cuentas, la administración recurre a informes anuales de gobierno en donde sintetiza las acciones realizadas y el monto ejercido. Por lo anterior, la integración de un enfoque evaluativo que trascienda la visión de “dar apoyos” y que permita una orientación a resultados requiere iniciar con mejoras en el diseño de los dos instrumentos contemplados en la normatividad municipal en materia de planeación: el plan municipal de desarrollo y los programas operativos anuales de la Dirección de Desarrollo Agropecuario y la Coordinación de Ecología.

Con respecto del registro de un uso de proceso, precondition que favorece el uso de evaluaciones con fines de aprendizaje, algunas de las evaluaciones con las que interactúan la OSC y los guardaparques de la Dirección de ANP les involucran directamente, partiendo del supuesto de que ellos, y no actores externos, son los principales usuarios de los hallazgos.

En el caso de los guardaparques, el ciclo de planeación-evaluación de su POA genera un espacio para reflexionar y hacer ajustes en las actividades con base en la experiencia que han adquirido. A su vez, la evaluación del programa de manejo vigente y las evaluaciones de la efectividad en el manejo les han servido para identificar brechas en los procedimientos que utilizan. En el caso de la OSC, la gestión de proyectos multianuales ha implicado evaluaciones intermedias con la participación de evaluadores externos. Estas han servido como un espacio de reflexión para sistematizar y reportar “lecciones aprendidas” que orientan las actividades futuras del proyecto.

7.3.2 Nivel organizacional

Las únicas organizaciones que cuentan con puestos en la escala local especialmente dedicados a las evaluaciones son la AM y la OSC. Si bien, en la figura de la contraloría municipal el ayuntamiento cuenta con un órgano desconcentrado responsable de evaluar el desempeño, en términos reales el área se enfoca en auditar el uso correcto de los recursos con la finalidad de prevenir eventos de corrupción. La revisión que hace de la entrega de los programas operativos anuales es superficial, pues no retroalimenta su diseño ni verifica su cumplimiento. Dada la asociación inevitable de la contraloría con la fiscalización, sería conveniente que el ayuntamiento tuviera una figura similar de combate a la corrupción, pero enfocada en procesos de mejora de la efectividad organizacional. Después de todo, la corrupción puede presentarse tanto en el desvío ilegítimo de recursos, como en la falta de capacidad para resolver problemas públicos. Esta instancia, responsable de evaluar los programas de desarrollo del municipio, podría trabajar con las diferentes áreas del ayuntamiento para identificar debilidades en el diseño de sus planes y crear consensos en torno de la reestructuración de su teoría del cambio. Además, si fueran funcionarios de carrera podrían mejorar con el tiempo sus capacidades para hacer y usar evaluaciones. Esto permitiría trascender una de las principales limitaciones de esta organización: la pérdida de conocimientos debido a la rotación de personal cuando cambian las administraciones municipales.

En cambio, la OSC se apoya en un especialista que asesora el diseño y operación de su sistema de evaluación. Su experiencia se concentra en el desarrollo de modelos lógicos, la valoración económica de ecosistemas, y el cálculo del retorno social de la inversión. En ese sentido, él ha sido responsable de integrar las evaluaciones a las propuestas de proyectos multianuales con donantes internacionales y de actualizar regularmente las evaluaciones internas que emplea la organización.

Con respecto de la disponibilidad de sistemas locales de gestión de información, procedimiento organizacional que permite dar seguimiento al trabajo realizado, la OSC y la dirección de ANP son las únicas que cuentan con este tipo de sistema de soporte, materializados en sistemas de información geográfica y formatos de registro. En contraste, el municipio y la universidad no cuentan con procedimientos para almacenar la información generada a través de la implementación de sus proyectos.

Sin embargo, un obstáculo que debilita este tipo de sistemas, reportado por todas las organizaciones, es la falta de recursos para dar seguimiento a los proyectos. En buena medida, la operación se centra en la ejecución de actividades, sin que se consideren acciones posteriores para verificar cambios y plantear adecuaciones. En general, esta es la lógica imperante en proyectos con una duración menor a un año, la cual afecta especialmente a las organizaciones gubernamentales, atadas a ejercicios fiscales. Fuera de los proyectos multianuales administrados por la OSC, en el presupuesto de los proyectos no existen montos asignados para el monitoreo per se, ni para llevar a cabo evaluaciones de la efectividad. Independientemente, las organizaciones manifestaron una sobrecarga de actividades en su personal que limita el tiempo disponible para recabar datos de seguimiento.

En un contexto en el que los recursos son exclusivamente para ejecutar proyectos, es evidente la brecha entre teoría y práctica cuando se habla del uso de evaluaciones para aprender. Por ello, si se desea mantener el propósito de manejar los proyectos a través de adaptaciones, es necesario incorporar procedimientos para sistematizar y utilizar el aprendizaje que adquiere el equipo interventor mientras realiza las acciones.

Dados los márgenes de incertidumbre en los que se desenvuelven los proyectos, lo anterior debería ser una prioridad tanto de donantes como de puestos superiores en la jerarquía organizacional, especialmente si desean que los recursos se destinen principalmente para la ejecución de actividades. Por ello, en un contexto carente de recursos para hacer evaluaciones, este tipo de procedimientos podrían sumarse a los mecanismos indirectos que funcionan como *proxy* para confiar en la obtención de resultados, como, por ejemplo, las evaluaciones *ex ante*. Sin embargo, de integrarse, hay que ser consciente de que es necesario otorgar flexibilidad operativa para que los equipos hagan ajustes tanto en las actividades como en el presupuesto.

De cualquier forma, es innegable que los proyectos con tiempos cortos de ejecución se benefician poco de este tipo de procedimientos. Ejemplo de lo anterior son los programas gubernamentales de subsidio que cambian año con año de beneficiarios. Por otro lado, por muy reflexivo que sea el equipo interventor, no se puede considerar al manejo adaptativo de los proyectos como una condición suficiente para conseguir impactos, sigue siendo necesario llevar a cabo evaluaciones *ex post* que determinen los niveles de logro.

En esa coyuntura, las universidades públicas podrían aportar mucho valor si construyeran alianzas con otras organizaciones que, junto a ellas, llevan a cabo proyectos en reservas de la biosfera. Por sus características, sería mucho más sencillo gestionar financiamiento para hacer investigaciones que evalúen la efectividad y retroalimenten el diseño de los proyectos de intervención. Desde un enfoque transdisciplinario, los investigadores podrían articular la demanda de conocimiento de las organizaciones y trasladarlo a preguntas científicas ¿Las obras de conservación de suelos están cumpliendo su función? ¿Ha incrementado la población de la especie amenazada? ¿Qué factores obstaculizan la adopción del pastoreo rotacional en la región? Sin embargo, más allá de la mera producción de conocimientos, es necesario que los investigadores se involucren en actividades de

intermediación que les permitan acompañar a las organizaciones tanto en el diseño como en la puesta en práctica de los proyectos.

7.3.3 Nivel institucional

La investigación identificó diferencias importantes en la influencia que ejerce el entorno institucional sobre los sistemas de evaluación de las organizaciones. En relación a la administración municipal, a pesar de que la Ley General de Contabilidad Gubernamental establece una serie de normas aplicables relacionadas con la programación basada en resultados, las cuales solicitan el uso de la metodología del Marco Lógico, en la práctica estos procedimientos no fueron adoptados por la administración.

La Entidad de Fiscalización del Estado es la instancia que verifica el cumplimiento de las normas, sin embargo, sus auditorías se han concentrado en torno del ejercicio del presupuesto. De esta forma, aunque existe una brecha de implementación en la consolidación de un Sistema de Evaluación del Desempeño que incluya un Presupuesto basado en Resultados, la administración municipal no se vio afectada.

A pesar de que esta situación quedó de manifiesto en una evaluación para mejorar la efectividad organizacional en la que, de forma voluntaria, participó el municipio, las recomendaciones no fueron retomadas. Por ello, es posible afirmar que la autonomía institucional del municipio le ha permitido mantener un déficit de rendición de cuentas en términos de los resultados generados a través de sus proyectos de desarrollo rural.

Institucionalizar una cultura evaluativa requiere, por supuesto, capacitaciones para que las autoridades municipales valoren su importancia, pero, para no quedar sujeta a la discrecionalidad de la administración en turno, sería importante contar con otro tipo de mecanismos que refuercen su aplicación. A nivel externo, por ejemplo, podría sugerirse a la Entidad de Fiscalización del Estado el no limitarse a la auditoría de aspectos financieros, podría incorporar también la revisión de una programación basada en resultados.

En un sentido semejante, pero a nivel interno, otro mecanismo podría ser la demanda ciudadana de estándares organizacionales que garanticen la toma de decisiones sobre bases racionales. Para este fin, se podría aprovechar la figura de los consejos municipales de participación ciudadana, los cuales están previstos en la ley orgánica municipal del Estado al que pertenece el municipio. Estos, entre otras, tienen la competencia de participar en los procesos de planeación y evaluación de programas municipales.

Además, dado que los municipios tienen la capacidad de dictar sus propias normas, se podría impulsar la emisión de un reglamento que detalle los procedimientos a seguir en materia de evaluación. Considerando lo anterior, sería relevante la realización de un proyecto de investigación aplicada que combine (1) la oferta de capacitaciones y (2) el desarrollo de un marco normativo con (3) un análisis de la efectividad de las estrategias de intermediación empleadas para promover cambios

en el sistema de evaluación del ayuntamiento.

En contraste, la OSC tiene una evidente profesionalización en materia de evaluación, la cual ha adquirido a lo largo de más de treinta años de concursar por fondos internacionales. Este caso refleja cómo la presión por adaptarse a un entorno que demanda capacidades de evaluación resultó en la integración de personal especializado y en la adopción de prácticas que se consideran un estándar en el campo del desarrollo, como el uso de modelos lógicos e indicadores de resultado.

La evidencia de su proceso de aprendizaje se expresa en la sofisticación de los proyectos a evaluar. Mientras que en el caso de la administración municipal, donde la rotación del personal es la norma y la gestión de recursos externos casi inexistente, las iniciativas de manejo de recursos naturales se basan en conceptos sencillos, como campañas de reforestación con el apoyo de jóvenes, capacitaciones en la separación de residuos, o la entrega de insumos para la producción; la OSC, en cambio, con base en la continuidad de puestos directivos, ha transitado de la educación ambiental y la reforestación de predios con fines de aprovechamiento forestal en los ochentas, a una estrategia diversificada que incluye herramientas de conservación basadas en el mercado y un programa de compra de tierras para conformar corredores ecológicos en reservas privadas. La obtención de financiamiento para dichos proyectos implicó interacciones con múltiples requerimientos evaluativos que, sumados, generaron una espiral de aprendizaje al interior de la organización.

Sin embargo, las oportunidades de aprendizaje no han provenido exclusivamente del uso instrumental de evaluaciones, sino de una mezcla de interacciones con otras organizaciones en premiaciones y conferencias. Ejemplo de ello fue la integración de la metodología del Retorno Social de la Inversión (SROI), el sistema de evaluación que más ha empleado la organización para comunicar sus resultados. Su adopción provino, no de los requerimientos de un donante, sino por elección de la propia organización luego de ser confrontada en un encuentro sobre la falta de evidencia de los cambios que aseguraba estar generando.

En el caso de la universidad pública, coincidiendo con lo señalado para varias universidades del país, los indicadores de evaluación académica incorporados en documentos de planeación, informes de resultados y procedimientos de contratación retoman los propuestos por la política científica federal, vinculando el impacto con la productividad académica.

Estos indicadores forman parte de la evaluación de distintos programas de gobierno que brindan fondos extraordinarios a la universidad, los cuales, entre otros aspectos miden la “capacidad académica” de su personal. La universidad ha manifestado la importancia de este tipo de convocatorias para recibir recursos adicionales, pues enfrenta un contexto de carestía presupuestal que limita su desarrollo⁸². Por ello, a través de sus procedimientos, la universidad ha transferido los criterios de evaluación de la organización en su conjunto hacia la evaluación de proyectos y su

⁸² La universidad destina el 100% del subsidio ordinario (federal y estatal) para pagar sueldos y prestaciones de su personal, y destina una buena cantidad de recursos propios para completar el pago de nóminas.

personal. Esto, a su vez, influye en el comportamiento de los académicos, pues determina qué actividades y resultados son valorados por la institución.

Sin embargo, existe una contradicción entre esos valores y la Responsabilidad Social Universitaria, modelo de gestión que se orienta hacia la medición y manejo de impactos organizacionales. Es interesante reflexionar en el margen de maniobra que tienen las Universidades Públicas Estatales para proponer esquemas distintos a los propuestos en políticas científicas que dotan de recursos a la organización en su conjunto y a sus investigadores a través de programas de pago por mérito. En definitiva, su autonomía institucional es un campo de acción para decidir si reproducen estructuras hegemónicas o si adoptan valores propios que conduzcan su comportamiento.

Finalmente, en relación a la CONANP es posible rastrear influencias a distintas escalas. A nivel internacional, la demanda externa para cumplir con la meta 11 de Aichi, la cual solicita aportar evidencia de que las áreas se manejan de forma efectiva, promovió la incorporación de instrumentos de evaluación de la efectividad organizacional como INDIMAPA e i-efectividad. Además de su uso para rendir cuentas y, potencialmente, para mejorar procesos, la adopción voluntaria de este tipo de evaluación se liga a la expectativa de que facilitará la gestión de fondos adicionales para la dependencia al cumplir con este requisito de convocatorias de financiamiento externo. La adición de recursos es, definitivamente, algo crucial ante la brecha presupuestal por la que atraviesa la CONANP.

A escala nacional, la dependencia precisa cumplir con diferentes lineamientos asociados a la gestión basada en resultados que emplea el gobierno federal. En primer lugar, precisa aportar información para las evaluaciones externas a programas de subsidio que supervisa el CONEVAL. Por su escala y propósito, el usuario potencial es el personal directivo de la dependencia, quien puede retomar la información para rediseñar los programas. Un aspecto interesante para valorar su influencia es que, a pesar de que las evaluaciones han observado, en repetidas ocasiones, que el indicador de propósito utilizado por el PROCODES no permite medir los efectos generados, este nunca ha sido modificado. Por otro lado, CONEVAL ha rechazado sistemáticamente las metodologías de evaluación de impacto propuestas por la CONANP, dado que estas no siguen el estándar dorado de emplear diseños experimentales para estimar los efectos. En este punto es evidente la distancia entre el enfoque de evaluación de CONEVAL y múltiples procedimientos de evaluación interna desarrollados por la CONANP, los cuales se centran en generar información a escala local para entender la dinámica del manejo de las áreas. Sin embargo, como ya se ha dicho, un desafío de la dependencia es la brecha de implementación de sus propios procesos evaluativos.

De forma adicional, tanto la Ley de Planeación como el Sistema de Evaluación del Desempeño previsto por la Ley Federal de Presupuesto y Responsabilidad Hacendaria, implican que la CONANP tiene la responsabilidad de comunicar el grado de cumplimiento de las metas establecidas en el PNANP, documento de planeación estratégica de la dependencia alineado al Plan Nacional de Desarrollo. Por ello, para facilitar la integración de la información a nivel de Dirección de ANP, la dependencia desarrolló el Sistema General de Programas Operativos Anuales, el

cual estandariza la información que las direcciones de ANP pueden reportar, con la finalidad de que sirva para alimentar los indicadores de resultado con los que se mide el desempeño global de la dependencia. En torno de este propósito, los guardaparques señalaron tensiones entre la pretensión de que ellos diseñen por sí mismos su POA en función de las problemáticas que observan, y solicitudes de que integren actividades alineadas con los objetivos estratégicos de la dependencia.

Si bien existen excepciones, el personal de las organizaciones refirió la predominancia de evaluaciones sumativas basadas en la entrega de resultados cuantificables predeterminados. En consecuencia, 1) el limitado margen para reportar fallas y resultados inesperados, y 2) el sesgo hacia el reporte de números por sobre análisis de procesos, son obstáculos importantes para el uso de evaluaciones formativas. Esta situación impacta también en la calidad de los procedimientos que se supone sirven para aprender, pues normalmente no se transparentan las técnicas a través de las que se busca promover un proceso reflexivo que conduzca a, por ejemplo, entender mejor el contexto de intervención, ajustar estrategias o identificar necesidades de capacitación en el equipo.

Un reflejo de la dominancia de los números por sobre las historias puede observarse en el uso que hace la OSC de la metodología SROI. Preponderantemente, la organización la emplea como una ventaja competitiva en la gestión de recursos al rendir cuentas sobre su impacto expresándolo en el lenguaje de tomadores de decisión que prefieren análisis costo/beneficio. Sin embargo, aunque la metodología original contempla el involucramiento recurrente de los participantes y la integración de indicadores cualitativos sensibles a cambios sociales, su análisis se ha simplificado para enfocarse en indicadores monetizables.

En ese sentido, es evidente la brecha entre su interés en metodologías para medir la captura de carbono, información que la organización requiere para ofertar compensaciones, y su interés en métricas de impactos sociales que trasciendan los basados en el ingreso. Así como se precisa rigor para validar protocolos locales de medición de captura de carbono, el establecimiento de impactos sociales no es tan simple como calcular el efecto multiplicador del dinero recibido o planificar reuniones entre beneficiarios y donantes.

Por ello, dado el carácter innovador de la organización, sería recomendable hacer más sofisticada su medición de impactos sociales, incorporando, por ejemplo, indicadores de bienestar (Woodhouse *et al.*, 2015). Entre otros aspectos, podría monitorear si los cambios en ingresos en verdad detonan mejoras en la calidad de vida, si sus proyectos promueven procesos de empoderamiento o de dependencia en los grupos beneficiarios, o la posible existencia de efectos no deseados por la forma en que se distribuyen los beneficios en las comunidades (Woodhouse *et al.*, 2018).

Finalmente, dada su experiencia y reconocimiento internacional, sería interesante que experimente con enfoques de evaluación centrados en la documentación de errores y desafíos por superar. El reporte de fallas empleado por la organización Ingenieros sin Fronteras es un ejemplo de metodologías de este tipo (EWB, 2017). Ese conocimiento sería útil, no solo para transparentar aún más su rendición de

cuentas, sino para estimular conversaciones con otras organizaciones locales en la búsqueda de soluciones a problemas complejos.

Además, en un futuro sería interesante estudiar el paradójico fenómeno de la promoción de reservas privadas dentro de una reserva de la biosfera, quizás como parte de una investigación más amplia sobre prioridades de financiamiento en el campo de la conservación. La preferencia de modelos excluyentes y la falta de medición de indicadores sociales podrían señalar el debilitamiento en la confianza de la efectividad de enfoques integrales de conservación.

Por su parte, como se ha mencionado para el caso de la CONANP, es importante que las evaluaciones de la efectividad en el manejo sirvan para que los guardaparques hagan cambios procedimentales, y no sólo para que la dependencia compruebe a través de un número el cumplimiento de la meta 11 de Aichi. Dado que es un proceso en marcha en la dirección de ANP, valdría la pena estudiar si estos tuvieron efectos sobre la toma de decisiones de los guardaparques o si se redujeron al reporte de índices de efectividad.

Más aún, sería importante identificar si los guardaparques interactúan con estas evaluaciones más preocupados por el número que arrojan sobre su desempeño, que por el espacio de reflexión que provocan para identificar mejoras en el manejo del ANP. Lo anterior es relevante dado que la investigación, como ya ha sido mencionado, identificó en torno de la evaluación del POA un comportamiento administrativo estratégico en los guardaparques, de forma que pudieran cumplir con las métricas de desempeño esperadas.

De forma similar, el sistema de evaluación de la universidad pública, al enfatizar métricas cuantitativas de productividad académica, corre el riesgo de desplazar la búsqueda de impactos a través del involucramiento de los investigadores en procesos de cambio, por una visión estratégica ligada al aumento en el número de publicaciones. Al asumir su responsabilidad social, la universidad necesita incorporar en sus procedimientos una visión de largo plazo que incentive tanto la generación de conocimiento como las actividades de intercambio a través de las que equipos interdisciplinarios de académicos colaboren en la co-creación de soluciones.

7.4 REFLEXIONES FINALES

La investigación estudió cuatro organizaciones que buscan generar valor público al implementar proyectos que impactan en el manejo de una reserva de la biosfera. ¿Las evaluaciones que realizan les permiten rendir cuentas con base en la detección de cambios en variables ambientales? ¿Les permiten juzgar la efectividad de sus estrategias con la finalidad de mejorarlas? Con pocas excepciones, los resultados probaron que las evaluaciones se reducían a verificar el cumplimiento de promesas preestablecidas, normalmente medidas a través de metas numéricas de corto plazo. Este hallazgo es común en la literatura sobre el tema, reforzando la

evidencia del dominio de evaluaciones sumativas en los campos de la conservación y el desarrollo.

Evidentemente, este enfoque de evaluación es insuficiente para cumplir con las expectativas de manejar los proyectos a través de adaptaciones en ANP. Para cambiar esta situación es necesario repensar qué valoramos en proyectos de esta naturaleza, pues el mérito basado solo en números, aunque nos parezca objetivo, coloca en una caja negra la dinámica del cambio y provoca distorsiones en el comportamiento de los evaluados.

En múltiples ámbitos de nuestra sociedad tenemos evidencia de estos errores en la forma de evaluar. Un ejemplo paradigmático se encuentra en el uso de calificaciones para valorar el desempeño educativo. En la década de los setenta, Susan Harter (1978) hizo un célebre experimento con niños para detectar la forma en que las calificaciones influían en sus decisiones al realizar anagramas con distintos grados de dificultad. Al estar ausentes las valoraciones numéricas, los niños elegían y disfrutaban anagramas que les empujaban al límite de sus habilidades; pero, al incluir calificaciones en la ecuación, los niños elegían consistentemente anagramas más sencillos, reportaban menos placer y verbalizaban más ansiedad. Es decir, cuando trabajaban por calificaciones, los niños bajaban su desempeño y se sentían peor.

La infancia y la adolescencia son etapas en las que definimos quiénes somos e identificamos dónde estamos parados. A través de la escuela internalizamos que la motivación principal de realizar una actividad (p. ej. poner atención en clase) es obtener recompensas (notas altas) y evitar sanciones (bajos puntajes). Por ello, desde niños reaccionamos a las valoraciones numéricas con miedo, estrés y comportamientos estratégicos como la evasión de tareas desafiantes o la memorización “bulímica” de los contenidos que vendrán en el examen.

Estas reglas del juego educativo no solo son inefectivas para promover el aprendizaje, también son contraproducentes. Por extensión, ¿qué podemos esperar cuando imbuimos la misma lógica en la evaluación de proyectos de desarrollo? Es común esperar que las evaluaciones se perciban como instrumentos de poder y control entre actores desiguales, las cuales, más que un medio colaborativo de aprendizaje, son un mecanismo para establecer niveles esperados de desempeño. En consecuencia, tal como sucede con los estudiantes, las organizaciones podrían enfocarse en el cumplimiento de los resultados de corto plazo impuestos por la administración de un proyecto, y no necesariamente en procesos difíciles de ejecutar y medir.

En el fondo, me parece que la raíz de nuestra relación problemática con las evaluaciones se asocia a un fenómeno propio de sociedades meritocráticas: la ansiedad por el estatus. Después de todo, una evaluación negativa es asociada por el individuo con la pérdida de credibilidad, respeto o aceptación; experiencias negativas que la mayor parte de las personas intentamos evitar (Donaldson *et al.*, 2002).

Alain de Botton ha explorado el tema de forma brillante en su libro *Status Anxiety* (2004). En primer lugar, dice, la preocupación por la posición que tenemos en la

sociedad deriva de una incertidumbre congénita sobre nuestro propio valor, motivo por el cual lo que otros piensan de nosotros juega un rol determinante en cómo nos percibimos. Por ello, dado que nuestro valor es captivo de los juicios de nuestro grupo, no nos sentimos avergonzados por fallar en sí, nos sentimos humillados cuando somos incapaces de cubrir expectativas sociales sobre lo que deberíamos haber conseguido.

El problema se agudiza, afirma Botton, en sociedades meritocráticas que colocan toda la responsabilidad sobre las cualidades del individuo al valorar su desempeño. En oposición, la ansiedad por el estatus es mucho menor en culturas que dan un lugar importante al azar, pues comprenden que en buena medida la fortuna de las personas depende de eventos sobre los que no tienen control. Sin embargo, el rol de la incertidumbre en la evaluación del mérito de las personas se ha erosionado en la era tecnocrática, la cual asigna al hombre un gran poder para prever y controlar su entorno. En el campo del desarrollo, esa racionalidad está detrás del supuesto de que el cambio es proyectable, y que por tanto, puede establecerse un contrato en el que se acuerden por adelantado los resultados a alcanzar.

Por ello, tan sumidos como estamos en discursos de productividad y control, no deseamos abandonar la sensación de seguridad, objetividad y precisión que nos da el uso exclusivo de *proxies* numéricos para medir el desempeño. Así, cuando enfrentamos preguntas difíciles, tendemos a sustituirlas por preguntas que pueden responderse fácilmente con números. Por ejemplo, para conocer el impacto social de la actividad científica en una reserva nos preguntamos, ¿a cuánto asciende la productividad académica? O, para saber si promovemos una ganadería sostenible, nos preguntamos, ¿cuántos talleres hemos dado y cuántos subsidios otorgamos?

En definitiva, es duro vivir en una cultura empapada de nociones de control en donde, como afirma Bauman, valemos lo que vale nuestro último éxito. Por ese motivo, tenemos muchos incentivos para simplificar y distorsionar los indicadores con los que se mide nuestro desempeño. Después de todo, si dependemos más de la opinión de otros que de resultados difíciles de medir, ¿por qué no crear la impresión de que las cosas marchan bien usando métricas numéricas? Eso explica la vigencia del proceso psicológico descrito por la ley de Campbell, y también la actualidad del aforismo del duque La Rochefoucauld: vivimos en una sociedad que con mucha mayor frecuencia recompensa signos del mérito que el mérito en sí.

Lamentablemente, estas ideas están muy arraigadas en nuestro sistema de evaluación. Transformarlo implica reconfigurar nuestros procesos de valoración, de forma que reconocer un error no se asocie a admitir que se es un fracaso. Al contrario, en el campo del desarrollo es necesario reinterpretar el aprendizaje a partir de fallas como un triunfo muy valioso. Por ello, sería conveniente que el personal de las organizaciones, en lugar de tener miedo por no cumplir las promesas hechas a donantes, fuera apoyado por estos para adaptar los proyectos, y por ende, rendir mejores cuentas a los beneficiarios de las intervenciones. En contextos complejos, el potencial que depositamos en la evaluación como mecanismo de mejora depende de que esto suceda.

En el fondo, para que se usen las evaluaciones para aprender, necesitamos un entorno normativo que promueva que el personal de las organizaciones adopte la dimensión interna de la rendición de cuentas, la cual implica ser responsable de lo que se ha hecho y comprometerse con los objetivos que se desea conseguir. Esa orientación ética, más que la solicitud externa de resultados, debería ser la base que impulse la búsqueda de respuestas para responder de mejor manera ante contextos cambiantes de intervención.

Los resultados de la investigación muestran que, si bien el entorno institucional ejerce una influencia importante, las organizaciones también tienen el poder de establecer qué es lo que cuenta. Por ello, existe la oportunidad de aprovechar su margen de autonomía para vincular las evaluaciones con la búsqueda activa de transformaciones que favorezcan a las personas y al ambiente. En particular, pienso en el potencial que tiene incidir en posiciones de liderazgo, como la dirección de una ANP, en la rectoría de una universidad o en la presidencia municipal. Estos puestos son responsables de alinear valores, recursos e incentivos para que los equipos de trabajo se orienten hacia resultados de largo plazo.

Dicho lo anterior, quisiera destacar tres aspectos que considero elementales para desarrollar las capacidades de evaluación de las organizaciones estudiadas. Juzgo que son elementos técnicos apropiables por líderes comprometidos, considerando los obstáculos que enfrentan, así como el margen que tienen para reformar su sistema de evaluación.

a) Valorar las victorias pequeñas.

Las cifras que usamos para medir productos, si bien se obtienen con facilidad, son un medio inadecuado para describir procesos de cambio. Por ello, para que el énfasis de la evaluación se ponga en el aprendizaje y la mejora, y no en juicios sumarios del desempeño, hay que saber qué sí y qué no debemos contar. Como bien dijo James Gustave Speth: dado que tendemos a obtener lo que medimos, empecemos a medir lo que queremos.

En ese sentido, es importante distinguir entre las victorias rápidas, asociadas a la entrega de productos y la atención de síntomas; y las victorias pequeñas, que miden resultados concretos de importancia moderada (Weick, 1984) porque causan efectos de corto y mediano plazo (Cuadro 31).

Cuadro 31. Diferencias entre victorias rápidas y pequeñas

Victorias rápidas	Victorias pequeñas
5 talleres realizados. Artículo en revista indizada.	Inicio de acciones. Seis meses después de la capacitación, el 30% de los agricultores adoptaron la tecnología. Establecimiento de consensos. Como resultado del diálogo multi-actor, las

1 hectárea reforestada empleando incentivos directos.	autoridades municipales de la región aprobaron un reglamento municipal para regular la pesca de langostinos de río.
---	---

Fuente: Elaboración propia

Las victorias pequeñas tienen las siguientes características: 1) van más allá de promesas o ideas creativas, 2) implican hacer visibles los primeros pasos del cambio que se desea obtener, al influir en valores, habilidades o decisiones de actores específicos y 3) permiten al equipo reflexionar sobre los mecanismos que permitieron, o no, obtener el resultado (Catrien y Dewulf, 2019).

En este contexto, en lugar de valorarse el desempeño a través de resultados medibles en el corto plazo, las evaluaciones deberían verificar los esfuerzos de adaptación de las organizaciones para conseguir victorias pequeñas, como el fortalecimiento de redes o el establecimiento de consensos (Cuadro 32). Además, una valoración justa del éxito de los proyectos tendría que considerar las relaciones de confianza que se fueron construyendo, la toma de riesgos y la creatividad que produjeron innovaciones útiles, así como las herramientas que fueron refinadas con el paso del tiempo. Si no se le da peso a los años de trabajar y esperar, aquellos en los que se dan dos pasos hacia adelante y un paso hacia atrás, entonces el éxito resultante no se entiende de forma correcta (Patel, 2007 en Guijt, 2008: 19).

Cuadro 32 Formas concretas de conceptualizar las victorias pequeñas

Tipo de cambio	Efecto esperado
Nuevos conocimientos	Es más probable que una persona actúe gracias a un cambio en su actitud o comprensión sobre un tema.
Inicio de acciones	Una persona o grupo inicia o modifica una actividad gracias a lo que se aprendió, practicó o realizó.
Establecimiento de consensos	Es más probable que un grupo con una agenda común actúe. Esa agenda se puede construir gracias a nuevos conocimientos que promuevan una comprensión conjunta o mejoren los medios para colaborar.
Desarrollo de habilidades	Una persona es más capaz de actuar porque tiene una habilidad nueva o fortalecida.
Fortalecimiento de redes	Es más probable que un grupo actúe debido a las nuevas relaciones que formó, a que desarrolló más afinidad, confianza o a la reducción del aislamiento en el que se encontraban sus integrantes.

Fuente: Kumar *et al.* (2018).

b) Planear como requisito para evaluar.

Antes de pensar en evaluar es necesario proyectar a largo plazo y paso por paso la ruta que se desea poner a prueba para atender un problema complejo. Las planeaciones estratégicas funcionan como una brújula para no perder la dirección de los objetivos que se persiguen. En relación a ellas, las evaluaciones formativas permiten tantear el terreno, hacer mapas fidedignos del camino recorrido, e independientemente de lo que se desee, indicar si se está extraviado o en ruta correcta.

Lamentablemente, con frecuencia esta obviedad no se cumple, pues se invierte poco tiempo para planear en las organizaciones. Esto puede ligarse tanto a la falta de conocimientos técnicos como a la atención exclusiva en productos entregables, pues se adopta una visión episódica de los proyectos y, al terminarlos, ya se está pensando en los siguientes, sin dar importancia a la continuidad de los procesos.

Hay muchos ejemplos de nuestra preferencia por planes con rutas cortas de implementación. La sencillez de su lógica nos da confianza y nos permite comunicar claramente sus ventajas para persuadir a otros. Sin embargo, este tipo de planeaciones, basadas en el supuesto de que todo saldrá bien, no comprenden la complejidad y latencia de los procesos de cambio.

Por ello, es fundamental que las organizaciones integren teorías del cambio que concatenen las victorias pequeñas que pretenden conseguir. Sin embargo, deben tener la libertad de tratarlas como un experimento a probar y no como un mapa autorreferencial que predice lo que necesariamente ocurrirá. Después de todo, no tiene sentido ajustar un mapa si solo se puede transitar por caminos preestablecidos.

c) Abrir el espacio para la reflexión.

Cuando hacen falta oportunidades de retroalimentación, uno puede practicar algo de forma recurrente y aun así no mejorar. Nunca podremos mejorar nuestra puntería si lanzamos dardos todo el día sin saber dónde está la diana. En ese sentido, el aprendizaje es más probable cuando las personas reciben retroalimentación clara después de intentar algo.

Por este motivo, la existencia de capacidades organizacionales para reflexionar sobre la práctica es tan importante como cualquier competencia técnica en materia de evaluación. Sin embargo, las habilidades para facilitar conversaciones que alimenten ciclos de aprendizaje iterativo suelen ser poco reconocidas. Eso ocurre a pesar de que todos hemos participado en múltiples reuniones que se sienten como una pérdida de tiempo o en diálogos que terminan en una lucha de egos. Con frecuencia, las organizaciones adolecen de facilitadores que les permitan tener reuniones efectivas basadas en conversaciones que sean significativas a los asistentes. En este contexto, es relevante capacitar al personal en técnicas para estimular el pensamiento evaluativo, tal como el diálogo apreciativo, el café pro-acción o el espacio abierto.

La emergencia de espacios seguros de reflexión requiere, por supuesto, construir lazos de confianza basados en el reconocimiento de la posibilidad de fallar. Desde

esta lógica, no te desempeñas mal si fallas, en cambio, te desempeñas mal si ocultas tus fallas porque eso impide identificar en qué información se apoyan los supuestos que condujeron al error.

Una vez trascendido lo anterior, el equipo puede tener ejercicios recurrentes de reflexión que le permitan hacer sentido de lo que ocurre con los proyectos para adaptarse en consecuencia. Cada intento por obtener una victoria pequeña, independientemente de si es exitosa o no, es una fuente potencial de aprendizaje porque los resultados proveen una retroalimentación de la efectividad de la estrategia.

La identificación de áreas de mejora puede beneficiarse de marcos conceptuales que permitan desmenuzar el tipo de falla que se tiene enfrente. De esa forma pueden estructurarse respuestas en función del nivel de responsabilidad e incertidumbre asociado a la falla (Cuadro 33).

Cuadro 33. Respuestas para diferentes tipos de falla.

		Tipo de falla	Ejemplo	Respuesta	
INCERTIDUMBRE - +	RESPONSABILIDAD +	Desviación del procedimiento acordado	El asesor técnico se puso de acuerdo con la mesa directiva del ejido para que firmaran los reportes sin que se hicieran las brechas corta fuego.	Uso de sanciones y sensibilización.	
		Desatención	Por descuido, María no se percató que la ventanilla del programa se cerraba el 2 de mayo. Aunque su A.C. elaboró un proyecto muy bueno, no pudieron ingresarlo para concursar por el financiamiento.	Desarrollo de capacidades.	
		Falta de habilidades	Es evidente que Ramón todavía no domina la técnica de inseminación artificial de ovinos. Por ese motivo, no hemos obtenido los resultados esperados.	Desarrollo de capacidades.	
	+	-	Estrategia inadecuada	Suponíamos que los jóvenes se interesarían en las actividades de conservación si utilizábamos las redes sociales para divulgar nuestra campaña de reforestación. Sin embargo solo le dieron <i>like</i> y ninguno llegó el día programado.	Revisión de teoría del cambio.
			Dificultad de la tarea	El proyecto busca unificar a los nueve municipios de la sierra, con la finalidad de que sus habitantes colaboren en la gestión adecuada de los residuos sólidos urbanos. Sin embargo, aún no hemos podido contar con el apoyo de todos los actores involucrados.	Revisión de teoría del cambio. Victorias más pequeñas.

		Prueba de hipótesis	El proyecto buscaba reforestar con acodos aéreos de <i>Prosopis Laevigata</i> , con la finalidad de tener individuos con una altura superior a un metro. Sin embargo, aún hace falta más investigación, pues con el sustrato que utilizamos solo el 10% generó raíces.	Brindar espacio para experimentar.
		Prueba exploratoria	El equipo interventor desconocía la existencia de liderazgos contrapuestos en la comunidad. Al convocar al taller eso se hizo evidente pues surgieron historias de conflictos pasados. Esa información condujo a un replanteamiento de las actividades.	Brindar espacio para ajustar los planes.

Fuente: Modificado de Edmondson (2011).

¿Cómo estimular la integración de estos elementos en las organizaciones estudiadas? Si se interioriza la lógica planteada por la tesis, es posible preguntarnos por el uso que tendrán los resultados generados por este proyecto de investigación. Por ello, para finalizar se presenta un cuadro que conecta las contribuciones empíricas con procesos de cambio en un plan de colaboración con la universidad pública (Cuadro 34). El objetivo es dar seguimiento a planes similares con la administración municipal y la dirección de ANP.

Cuadro 34. Plan de colaboración con la universidad pública estudiada.

Actores clave	Estrategia de intermediación	Tipo de cambio	Cambio esperado	Indicador
Investigadores	Curso de formación docente	Desarrollo de habilidades	Los investigadores aprenden a diseñar rutas de generación de impacto que incorporan estrategias de intermediación.	# de investigadores participantes que integraron teorías del cambio y estrategias de intermediación en el diseño de sus proyectos.
Rectoría de la Universidad Dirección de Investigación	Presentación de informe	Establecimiento de consensos	La administración de rectoría reconoce públicamente el valor de trascender métricas de productividad académica y firma la Declaración de San Francisco sobre evaluación de la investigación.	Declaración firmada. # de investigadores que conocen la declaración a los seis meses de haberse firmado.
Investigadores Dirección de Vinculación Dirección de Investigación	Maratón de escritura	Inicio de acciones	Los responsables de proyectos en la reserva aprenden a elaborar un reporte de fallas en el que sistematizan errores y aprendizajes. Los investigadores valoran la herramienta como mecanismo de aprendizaje y rendición de cuentas.	Publicación bianual de reporte de fallas.
Investigadores Dirección de Investigación	Premio “Ciencia con Impacto”	Fortalecimiento de redes	La universidad reconoce y promueve investigaciones aplicadas que son capaces de probar la generación de valor público. Con la convocatoria, la universidad comunica objetivos deseables como la construcción de alianzas, el enfoque hacia grupos vulnerables y la medición de impactos distintos a métricas de productividad.	Libro publicado con proyectos seleccionados.

Actores clave	Estrategia de intermediación	Tipo de cambio	Cambio esperado	Indicador
Direcciones de Facultades Investigadores	Conversatorios sobre evaluación de investigaciones	Establecimiento de consensos	Al reconocer la importancia de trascender el uso de métricas de productividad académica, los investigadores identifican desafíos, áreas de oportunidad y mejores prácticas en la evaluación de proyectos de las disciplinas pertenecientes a la Facultad.	# de investigadores participantes que pueden referir dos problemas asociados al uso de métricas de productividad académica a los 3 meses del encuentro. Documento con indicadores <i>ad hoc</i> para la evaluación del impacto de los proyectos de investigación y vinculación de la Facultad.
Rectoría de la Universidad Dirección de Investigación	Convocatorias de financiamiento focalizado	Inicio de acciones	A través de las convocatorias, la universidad promueve la atención de problemas específicos, la integración de rutas de generación de impacto (<i>ex ante</i>) y el reporte de indicadores de interacciones productivas (<i>ex post</i>).	Convocatorias modificadas Reporte de la pluralidad de impactos generados por los proyectos financiados.
Investigadores Consejo Universitario	Ajustes en reglas de ingreso y promoción académica	Inicio de acciones	A través del reglamento, la universidad incentiva que los investigadores mantengan interacciones productivas con otros actores pues solicita evidencias de uso de conocimientos y reconocimiento público.	Reglamento modificado. Incremento en el número de investigadores que reportan ese tipo de evidencia. Base de datos que georreferencia la pluralidad de impactos generados por la universidad a nivel estatal.

Fuente: Elaboración propia

Referencias

- Agudo, A., 2012. Consultorías (pos)modernas. La mímica del positivismo y la construcción del conocimiento en la evaluación de programas sociales. *Estudios Sociológicos*, 30 (88), 45-86.
- 2015. Una etnografía de la administración de la pobreza. La producción social de los programas de desarrollo. Ciudad de México: Universidad Iberoamericana.
- Alix J., Sims, K., y Yañez, P., 2014. Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's payments for ecosystem services program. *American Economic Journal: Economic Policy*, 7(4), 1-40.
- Alkin, M., 1985. *Organising for evaluation use, a handbook for administrators*. UCLA, Los Angeles: Center for the Study of Evaluation.
- Alkin, M. y Christie, C., 2004. An Evaluation Theory Tree. En Alkin, M. (ed.) *Evaluation Roots-Tracing Theorists' Views and Influences*. Thousand Oaks, Estados Unidos: Sage.
- Alkin, M. y King, J., 2016. The historical development of evaluation use. *American Journal of Evaluation*, 37 (4), 568-579.
- Alkin, M., y King, J., 2017. Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434-450.
- Alonso, G., 2014. Conservation Science Policies Versus Scientific Practice: Evidence from a Mexican Biosphere Reserve. *Human ecology review*, 20 (2), 3-29.
- Alvarez, S., Douthwaite, B., Graham T., Mackay, R., Cordoba, D. y Tehelen, K., 2010. Participatory Impact Pathways Analysis: a practical method for project planning and evaluation. *Development in Practice*, 20, 946-958.
- Andam, K., Ferraro, P., Pfaff, A., Sanchez, G., y Robalino, J., 2008. Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the national academy of sciences*, 105 (42), 16089-16094.
- Argyris, C., 1994. Initiating change that perseveres. *Journal of Public Administration Research and Theory: J-PART*, 4(3), 343-355.
- Argyris, C. y Schön, D., 1978. *Organizational Learning: A Theory of Action Perspective*. AddisonWesley, Massachusetts, 344 p.
- Argyris, C., y Schon, D., 1996. *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley
- Asian Development Bank, 2008. *Auditing the lessons architecture*. Philippines: Asian Development Bank.
- Baird, T., 2014. Conservation and unscripted development: proximity to park associated with development and financial diversity. *Ecol. Soc.* 19(1), 4.
- Banco Mundial, 2005. Política operacional 4.10 Pueblos Indígenas. Recuperado de <http://siteresources.worldbank.org/OPSMANUAL/Resources/2103841170795590012/OP4.1>

- Banco Mundial, 2015. World development report 2015: mind, society, and behavior. Washington DC: World Bank.
- Barnett, C. y Munslow, T., 2014. Workshop Report: Framing Ethics in Impact Evaluation: Where Are We? Which Route Should We Take? IDS Evidence Report 98. Inglaterra: IDS.
- Batliwala, S. y Pittman, A., 2010. Capturing Change in Women's Realities. A Critical Overview of Current Monitoring and Evaluation Frameworks and Approaches. Toronto: AWID.
- Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P., y Wunder, S., 2016. Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9(1), 58-64.
- Befani, B., 2012. Models of Causality and Causal Inference. En Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., y Befani, B., DFID Working Paper 38. Broadening the range of designs and methods for impact evaluations. Londres: DFID.
- Befani B., Barnett, C. y Stern, E., 2014. Introduction – Rethinking Impact Evaluation for Development. *IDS Bulletin*, 45 (6), 1–5.
- Behn, R., 2002. The psychological barriers to performance management: Or why isn't everyone jumping on the performance-management bandwagon? *Public Performance & Management Review*, 26(1), 5–25.
- Bell, S. v Aadleton. P.. 2012. Integrating ethnographic principles in NGO monitoring and impact evaluation. *Journal of International Development*, 24 (6), pp. 795-807.
- Bennett, N. J., Roth, R., Klain, S. C., Chan, K., Christie, P., Clark, D. A., Cullman, G., Curran, D., Durbin, T. J., Epstein, G., Greenberg, A., Nelson, M. P., Sandlos, J., Stedman, R., Teel, T. L., Thomas, R., Verissimo, D., & Wyborn, C., 2017. Conservation social science: Understanding and integrating human dimensions to improve conservation. *Biological Conservation*, 205, 93–108. <https://doi.org/10.1016/j.biocon.2016.10.006>
- Benoit, P., Colinet, L., Gaunand, A., Lemarié, S. y Matt, M., 2016. Agricultural research impact assessment: issues, methods and challenges. OECD Food, Agriculture and Fisheries Papers, No. 98. Paris: OECD Publishing.
- Bensing, J., Andeweg, R., Franses, Ph., Meyer, B., Prins, C., y Schuyt K., 2013. Towards a Framework for the Quality Assessment of Social Science Research. Amsterdam, Holanda: Royal Netherlands Academy of Arts and Sciences.
- Bentley, J., 2019. Decreasing operational distortion and surrogation through narrative reporting. *The Accounting Review*, 94(3), 27-55.
- Bertzky, M., 2009. Mind the gap: Information gaps and bridging options in assessing in-situ conservation achievements. Tesis doctoral, Ernst-Moritz-Arndt University Greifswald, Alemania.
- Bickman, L., 1987. The Function of Program Theory. En Bickman, L. (ed.), *Using Program Theory in Evaluation*. New Directions for Evaluation, no. 33. San Francisco: Jossey-Bass.
- Birnbaum, R., 2001. Management fads in higher education: where they come from, what they do, why

- they fail. San Francisco, CA: Jossey-Bass.
- Boege E., 2008. El Patrimonio Biocultural de los Pueblos Indígenas de México: Hacia la Conservación in situ de la Biodiversidad y Agrodiversidad en los Territorios Indígenas. Instituto Nacional de Antropología e Historia, Comisión Nacional para el Desarrollo de los Pueblos Indígenas México, D.F.
- Bongaarts, J., 2019. IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on biodiversity and ecosystem services. *Population and Development Review*, 45 (3), 680–681.
- Botcheva, L., White, C., y Huffman, L., 2002. Learning Culture and Outcomes Measurement Practices in Community Agencies. *American Journal of Evaluation*, 23(4), 421–434.
- Bourdieu, P., 1990. *The Logic of Practice*. Oxford: Blackwell Publishers.
- Bourgeois, I. y Cousins, B., 2013. Understanding Dimensions of Organizational Evaluation Capacity. *American Journal of Evaluation*, 00 (0), pp. 1-21.
- Bovens M., Schillemans T. y Hart P., 2008. Does public accountability work? An assessment tod. *Public Administration*, 86 (1), 225-242.
- Brandon, P., y Singh, J., 2009. The strengths of the methodological warrants for the findings on research on program evaluation use. *American Journal of Evaluation*, 30, 123–157.
- Brockington, D. y Wilkie, D., 2015. Protected areas and poverty. *Phil. Trans. R. Soc. B*, 370: 20140271.
- Brown, J. e Issacs. D.. 2005. *World café: Shaping our futures through conversations that matter*. San Francisco: Berrett-Koehler.
- Britton, B., 2005. *NGO and Organisational learning*. Praaxis Note nº3. Londres: INTRAC.
- Buckley J., Archibald, T, Hargraves, M. y Trochim, W., 2015. Defining and Teaching Evaluative Thinking: Insights From Research on Critical Thinking. *American Journal of Evaluation*, 36 (3), 375-388.
- Byrne, A. y Robin, V., 2004. *Learning in partnerships*. Briefing paper. Londres: BOND/Exchange, Londres.
- Cagan, R.. 2013. The San Francisco Declaration on Research Assessment. *Dis Models Mech* 6, 869-870, doi:10.1242/dmm.012955
- Camfield, L., Duvendack, M. y Palmer, R., 2014. Things You Wanted to Know about Bias in Evaluations but Never Dared to Think. *IDS Bulletin*, 45(6): 49–64.
- Campbell, D., 1976. *Assessing the Impact of Planned Social Change*. , Hanover New Hampshire: The Public Affairs Center, Dartmouth College.
- Campbell, D. y Stanley, J., 1978. *Diseños experimentales y cuasiexperimentales en la investigación social*. Buenos Aires: Amorrourtu Editores.
- Cardozo, M., 2015. Capítulo XXI. Revisión de la literatura. En Cardozo, M. (coord.), *Institucionalización de procesos de evaluación. Calidad y utilización de sus resultados*. Ciudad de México: Universidad Autónoma Metropolitana.

- Carranza, T., Manica, A., Kapos, V., Balmford, A., 2014. Mismatches between conservation outcomes and management evaluation in protected areas: A case study in the Brazilian Cerrado. *Biol. Conserv.* 173, 10-16.
- Castree N., Demeritt, D. y Liverman, D., 2009. Introduction: Making sense of environmental geography. En Castree, N., Demeritt, D., Liverman, D. y Rhoads, B. (eds.) *A companion to environmental geography*, Wiley-Blackwell, Singapur.
- Catsambas, T., 2016, Agosto. Reframing evaluation through Appreciative inquiry, en *Workshop in evaluation and applied research methods*, Claremont Evaluation Center.
- Cejudo, G., 2011. Sobre el uso de las evaluaciones. De las recomendaciones a las acciones: El uso de la evaluación en la mejora del desempeño del gobierno. En Cejudo, G. y Maldonado, C. (eds.), *De las recomendaciones a las acciones: La experiencia del premio 2011 Programas Federales Comprometidos con el Proceso de Evaluación*. México: CIDE-SFP.
- Cejudo, G. y Abarca, F., 2016. ¿Cómo se usan las evaluaciones? Un análisis de cómo los programas reaccionan a lo que recomiendan las evaluaciones. México, manuscrito.
- Cejudo, G. y Maldonado, C., 2011. *De las recomendaciones a las acciones: La experiencia del premio 2011 Programas Federales Comprometidos con el Proceso de Evaluación*. México: CIDE-SFP.
- Centre for Development Impact, 2015. *Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?* CDI Practice Paper 10
- Chacón, G. y Garduño, J., 2017, ¿Cuánto invierte el gobierno en la protección de nuestro hogar...nuestro patrimonio natural?, FUNDAR, recuperado de <http://noroosteenlamira.org.mx/wp-content/uploads/2017/03/infografia-APN-web.pdf>
- Chaplowe, S., y Cousins, B., 2016. *Monitoring and Evaluation Training: A Systematic Approach*, Sage Publications.
- Chavarro D. y Rafols I., 03 de enero de 2018. La evaluación de la investigación basada en revistas margina a regiones como América Latina y sus temas más relevantes. [Mensaje en un blog]. *Impact of Social Sciences*. Recuperado de <http://blogs.lse.ac.uk/latamcaribbean/2018/01/03/la-evaluacion-de-la-investigacion-basada-en-revistas-margina-a-regiones-como-america-latina-y-sus-temas-mas-relevantes/>
- Chelimsky, E., 1997. The coming transformation in evaluation. En Chelimsky, E. y Shadish, W., (eds.), *Evaluation for the 21st Century. A Handbook*. SAGE Publications, Londres.
- Chelimsky, E., 2015. A strategy for improving the use of evaluation findings in policy. En Christie, C. y Vo, A., (eds.) *Evaluation use and decision making in society. A tribute to Marvin Alkin*. Charlotte, Estados Unidos: Information Age Publishing.
- Christie, C., 2007. Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation*, 28, 8–25.
- Church, I. y Samuelson, P., 2017. *Intellectual Humility: An Introduction to the Philosophy and Science*, Bloomsbury Academic.
- Clemente, A., 2019, febrero 5. Conahcyt desaparecerá fondos mixtos para estados y municipios. El Financiero, <https://www.elfinanciero.com.mx/nacional/conacyt-desaparecera-fondos-mixtos->

para-estados-y-municipios.

Clements, T., Suon, S., Wilkie, D., Milner, G., 2014. Impacts of protected areas on local livelihoods in Cambodia. *World Development*, 64, S125-S134.

Climate, Community and Biodiversity Alliance (CCBA), 2013. Estándares de clima, comunidad y biodiversidad. CCBA, Arlington. Disponible en: http://verra.org/wpcontent/uploads/2016/05/CCB_Standards_Third_Edition_December_2013_Spanish.pdf; consultado: agosto de 2017

Cohn, A., Newton, P., Gil, J. D., Kuhl, L., Samberg, L., Ricciardi, V., y Northrop, S., 2017. Smallholder agriculture and climate change. *Annual Review of Environment and Resources*, 42, 347-375.

Comisión para la Cooperación Ambiental (2011). Guía para la elaboración de fichas de evaluación ecológica en áreas marinas protegidas de América del Norte. Montreal: CEC. Recuperado de <http://www3.cec.org/islandora/en/item/4184-guide-ecological-scorecards-marine-protected-areas-in-north-america-es.pdf>

CONABIO, 2007. Análisis de vacíos y omisiones en conservación de la biodiversidad terrestre de México: espacios y especies. México: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Comisión Nacional de Áreas Naturales Protegidas, The Nature Conservancy, Programa México, Pronatura, A.C., Facultad de Ciencias Forestales, Universidad Autónoma de Nuevo León.

CONANP (2010a), Evolución del Sistema de Información, Monitoreo y Evaluación para la Conservación, SEMARNAT, México.

----- (2010b), Revisión y evaluación de medio término sobre la Ejecución del programa nacional de áreas Naturales protegidas 2007-2012, SEMARNAT, México.

----- (2014a), Estrategia hacia 2040: una orientación para la conservación de las áreas naturales protegidas de México, SEMARNAT, México.

----- (2014b). Manual de organización general. México: SEMARNAT, 119 pp.

----- (2014c). Programa Nacional de Áreas Naturales Protegidas (2014-2018). México: SEMARNAT. Recuperado de <https://www.gob.mx/conanp/documentos/programa-nacional-de-areas-naturales-protegidas>.

----- (2016), Estrategia para la Evaluación de la Efectividad en la gestión de las Áreas Naturales Protegidas marinas y terrestres de México, SEMARNAT, México.

----- (2017a). Programa Nacional de Áreas Naturales Protegidas 2014 – 2018. Revisión y Evaluación de Medio Término. México: SEMARNAT, 19 pp. Comisión Nacional de Áreas Naturales Protegidas

----- (2017b). Sistema permanente de evaluación de la efectividad del manejo de las áreas naturales protegidas de México. México: SEMARNAT, 14 pp.

CONANP-GIZ, 2016. Áreas Naturales Protegidas. Región Centro y Eje Neovolcánico. México: Comisión Nacional de Áreas Naturales Protegidas, Secretaría de Medio Ambiente y Recursos Naturales.

CONANP-GIZ, 2018a. Prontuario Estadístico y Geográfico de las Áreas Naturales Protegidas de la

Región Centro y Eje Neovolcánico. México: Comisión Nacional de Áreas Naturales Protegidas, Secretaría de Medio Ambiente y Recursos Naturales.

CONANP-GIZ, 2018b. Lecciones aprendidas del proyecto conservación de la Biodiversidad en el eje neovolcánico 2014-2018. México: Comisión Nacional de Áreas Naturales Protegidas, Secretaría de Medio Ambiente y Recursos Naturales.

CONEVAL, 2016. Diagnóstico de elementos de monitoreo y evaluación en municipios capitales. México: CONEVAL

----- 2007. Guion del Análisis de Factibilidad para llevar a cabo una Evaluación de Impacto. México: CONEVAL

Conservation Measures Partnership, 2007. Open standards for the practice of conservation. Recuperado de: http://www.conservationmeasures.org/wpcontent/uploads/2010/04/CMP_Open_Standards_Version_2.0.pdf.

Cousins, B. y Leithwood, K., 1986. Current empirical research in evaluation utilization. *Review of Educational Research*, 56, pp. 331–364.

Cousins, B., Goh, S. y Clark, S., 2004. Integrating evaluative inquiry into the organizational culture: a review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, 19 (2), pp. 99-141.

Cousins, B., Amo, C., Bourgeois, I., Chouinard, J., Goh, S., y Lahey, R., 2008. Organizational capacity to do and use evaluation: results of a pan-canadian survey of evaluators, *The Canadian Journal of Program Evaluation*, 23 (3), pp. 1–35.

Cousins, B., Goh, S., Elliott, C., Aubry, T. y Gilbert, N., 2014. Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, 44, pp. 1-13.

Cousins B., Hay, K. y Chouinard, C., 2015. The third perspective. Uniting accountability and learning within an evaluation framework that takes a moral-political stance. En Christie, C. y Vo, A. (eds.). *Evaluation use and decision making in society. A tribute to Marvin Alkin*. Charlotte, Estados Unidos: Information Age Publishing.

Craig, L., Sanjayan, M., Blockhus, J., Kontoleon, A. y Larsen, N., 2010. Does conserving Biodiversity Work to reduce Poverty? A state of knowledge review. The Nature Conservancy, IIED.

Craigie, I., Barnes, M., Geldmann, J. y Woodley, S., 2015. International funding agencies: potential leaders of impact evaluation in protected areas?. *Phil. Trans. R. Soc. B*, 370: 20140283.

Cuervo, A. y Monroy, O., 2012. Distribución potencial del jaguar *Panthera onca* (Carnivora: Felidae) en Guerrero, México: persistencia de zonas para su conservación. *Revista de Biología Tropical*, 60(3), 1357–1367.

Cupitt, S., 2009. A guide to social return on investment. Office of the third sector. Londres: Cabinet Office.

Czarniawska, B., 2012. Organization theory meets anthropology: A story of an encounter. *Journal of Business Anthropology*, 1(1), 118-140.

- Darling, M., 2012. A compass in the woods: learning through grantmaking to improve impact, Fourth Quadrant Partners.
- David, T., s.f. Supporting Organizational Learning: From the Inside or the Outside. Community Clinics Initiative. Tides Network. Recuperado de: http://www.tdavid.net/uploads/3/8/9/1/38919813/inside_or_outside.pdf
- Dixon-Woods, M., y Bosk, C. L., 2011. Defending Rights or Defending Privileges?: Rethinking the ethics of research in public service organizations. *Public Management Review*, 13(2), 257–272. <https://doi.org/10.1080/14719037.2010.532966>
- De Botton, A., 2004. Status Anxiety. Edición Kindle.
- De la Rosa, M. 2016. Evaluando la eficacia de un área protegida costera ante el caambio del uso del suelo; la Reserva de la Biosfera Pantanos de Centla, México. Tesis de maestría. El Colegio de la Frontera Sur, 48 p.
- Department for International Development, 2011. Guidance on Using the Revised Logical Framework, How to note. A DFID practice paper. Department for International Development.
- Devi, S.. 2005. Ethics and development: Some concerns with David Mosse's Cultivating development. *Anthropology Today*, 21(6), 17–19.
- DiMaggio. P. v Powell. W.. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 147-160.
- Donaldson, S., Gooler, L. y Scriven, M., 2002. Strategies for Managing Evaluation Anxiety: Toward a Psychology of Program Evaluation, *American Journal of Evaluation*, 23 (3): 261–273.
- Donaldson, S. y Azzam, T. 2015. Tending the garden of evaluation theory. En Christie, C. y Vo, A. (eds.), *Evaluation use and decision making in society. A tribute to Marvin Alkin*. Charlotte, Estados Unidos: Information Age Publishing.
- Doolittle, A., 2007. Fortress conservation. En Robbins, P. (Ed.), *Encyclopedia of environment and society*. SAGE Publications. pp. 705.
- Earl, S., Carden, F., y Smutylo, T., 2010. Outcome Mapping: Building Learning and Reflection into Development Programs. Ottawa, IDRC.
- Ebrahim, A.. 2003. Making sense of accountability: Conceptual perspectives for northern and southern nonprofits. *Nonprofit Management and Leadership*, 14(2), 191-212.
- Ebrahim, A., 2005. Accountability Myopia: Losing sight of organizational learning, *Nonprofit and Voluntary Sector Quarterly*, 34 (1): 56-87.
- Edmondson, A., 2011. Strategies for learning from failure. *Harvard Business Review* 89, 48–55.
- Edwards, M., y Hulme, D., 1996. Too close for comfort? The impact of official aid on nongovernmental organizations. *World Development*, 24(6), 961-973.
- Ekboir, J., 2003. Why impact analysis should not be used for research evaluation and what the alternatives are. *Agricultural systems*, 78 (2), 166-184.
- 2016. Marco conceptual para el análisis institucional del INTA. Documento de trabajo del

Centro de Investigación en Ciencias Políticas, Económicas y Sociales, Buenos Aires: INTA.

- Ekboir, J., Dutrénit, G., Martínez, G., Torres, A., y Vera, A., 2009. Successful Organizational Learning in the Management of Agricultural Research and Innovation The Mexican Produce Foundations. IFPRI Research Report No. 162. Washington, DC: International Food Policy Research Institute.
- Ekboir, J., Blundo, G., y Sette, C., 2017. Knowing what research organizations actually do, with whom, where, how and for what purpose: Monitoring research portfolios and collaborations. *Evaluation and Program Planning*, 61, 64-75.
- Engineers Without Borders (EWB), 2017. 2017 failure report. EWB. Disponible en: https://www.ewb.ca/wp-content/uploads/2018/08/EWB_FAILURE-REPORT_EN_03-08-2018-pages.pdf
- Ervin, J., 2003. Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) Methodology Gland, Suiza: WWF.
- Esteves, A., Franks, D. y Vanclay, F., 2012. Social impact assessment: the state of the Art. *Impact Assessment and Project Appraisal*, 30(1), pp. 34-42.
- Eyben, R., 2008. Power, Mutual Accountability and Responsibility in the Practice of International Aid: A Relational Approach. *IDS working paper*, 305.
- 2009. Hovering on the Threshold: Challenges and Opportunities for Critical and Reflexive Ethnographic Research in Support of International Aid Practice. En Hagberg y Widmark (eds.), *Ethnographic Practice and Public Aid. Methods and Meanings in Development Cooperation*. Suecia: Uppsala University.
- 2013. Uncovering the Politics of 'Evidence' and 'Results'. A Framing Paper for Development Practitioners. Brighton: Institute of Development Studies.
- 2014. *International Aid and the Making of a Better World: Reflexive Practice (Rethinking Development)*. Taylor and Francis, Edición Kindle.
- 2015. Uncovering the politics of evidence and results. En Eyben, R., Guijt, I., Roche, C. y Shutt, C. (eds.), *The Politics of Evidence and Results in International Development: Playing the game to change the rules?*, Practical Action Publishing, Edición Kindle
- Eyben, R., León R. y Hossain N., 2007. Participatory Action Research into Donor-Recipient Relations: A Case Study. *Development in Practice*, 17 (2), 167-178.
- FAO, s.f.. REDD+ Reducing Emissions from Deforestation and Forest Degradation. Recuperado de: <http://www.fao.org/redd/overview/en/>
- Feinstein, O., 2002. Use of Evaluations and the Evaluation of their Use, *Evaluation*, 8 (4), pp. 433-439.
- Ferraro, P. y Pressey R., 2015. Measuring the Measuring the difference made by conservation initiatives: protected areas and their environmental and social impacts. *Phil. Trans. R. Soc. B*, 370: 20140270.
- Fetterman, D., 1996. Empowerment evaluation: An introduction to theory and practice. En Fetterman, D., Kaftarian, S. y Wandersman, A. (eds.), *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage

- Figuroa, F. y Durand, L., 2011. ¿Nos importan las cuestiones sociales? El contexto social en los programas de manejo de las reservas de la biósfera en México. Ponencia presentada en el quinto congreso Iberoamericano sobre Desarrollo y Ambiente de REDIBEC, 12-14 de septiembre, Santa Fe, Argentina.
- Forss, K., Rebien, C., y Carlson, J., 2002. Process use of evaluations. *Evaluation*, 8(1), 29–45.
- Friedman, R., Law, E., Bennett, N., Ives, C., Thorn, J., y Wilson, K., 2018. How just and just how? A systematic review of social equity in conservation research. *Environ. Res. Lett.* 13, 053001
- Funnel, S. y Rogers, P., 2011. Purposeful program theory: effective use of theories of change and logic models. Estados Unidos: Jossey Bass.
- Galaz, J., Padilla L., Gil M. y Sevilla J., 2008. Los dilemas del profesorado en la educación superior mexicana. *Calidad en la Educación*, 28, 53-69.
- Galaz, J., y Gil, M., 2013. The impact of merit-pay systems on the work and attitudes of Mexican academics. *Higher Education*, 66 (3), 357–374.
- Game, E., Bremer, L., Calvache, A., Moreno, P., Vargas, A., Rivera, B., y Rodriguez, L., 2018. Fuzzy models to inform social and environmental indicator selection for conservation impact monitoring: Fuzzv models for indicator selection. *Conservation Letters*, 11(1), e12338. <https://doi.org/10.1111/conl.12338>
- Gannon, P., Dubois, G., Dudley, N., *et al.*, 2019. Editorial essay: an update on progress towards aichi biodiversity target 11. *Parks*, 25, 2, 7-18.
- Geldmann, J., Barnes, M., Coad, L., Craigie, I., Hockings, M., y Burgess, N., 2013. Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation*, 161, 230-238.
- Gertler, P., Martínez, S., Premand, P., Rawlings, L. y Vermeersch, C., 2017. La evaluación de impacto en la práctica. Washington, DC: Banco Interamericano de Desarrollo y Banco Mundial.
- Giger, M., 2000. Evitando la trampa: Más allá del empleo de incentivos directos: Una sistematización de experiencias con el empleo de incentivos en proyectos de gestión sostenible de la tierra. Berna: Centro para el Desarrollo y el Medio Ambiente, Instituto de Geografía, Universidad de Berna.
- Gil, M., 2014. Los nuevos escenarios de la educación en México y el papel de las revistas científicas especializadas. *Perfiles educativos*, 36, 196-203.
- Gil, M. y Contreras L., 2017. El Sistema Nacional de Investigadores: ¿espejo y modelo? *Revista de la Educación Superior*, 46 (184), 1–19.
- Glover, D., Sumberg, J., Ton, G., Andersson, J., y Badstue, L., 2019. Rethinking technological change in smallholder agriculture. *Outlook on Agriculture*, 48(3), 169-180.
- González, A., 2013. Enfoque conceptual: El qué, porqué y para qué de la GpRD. En CIDE-Centro CLEAR para América Latina, Fortaleciendo la Gestión para Resultados en el Desarrollo en México: Oportunidades y Desafíos. México: CIDE-Centro CLEAR para América Latina, pp. 11-20
- González, A. y Cerdán, A. 2008. Marco teórico y analítico de la investigación. En González A. (Coord.)

¿Gobernar por resultados? Implicaciones de la política de evaluación del desempeño del gobierno mexicano. México: Gestión social y cooperación.

- González, J., 2016. Por qué se usa la información de seguimiento y evaluación (SyE) de políticas públicas, Tesis de Doctorado, Centro de Investigación y Docencia Económicas, México.
- Graham, B., 2006. *The Intelligent Investor: The Definitive Book on Value Investing*. Nueva York: Harper.
- Grammont, H., 2010. La evolución de la producción agropecuaria en el campo mexicano: concentración productiva, pobreza y pluriactividad. *Andamios*, 7 (13), 85-117.
- Grantmakers for effective organizations, 2016. Shaping culture through key moments. Recuperado de: <https://www.geofunders.org/resources/shaping-culture-through-key-moments-708>
- Guba, E. y Lincoln, Y., 2002. Paradigmas en competencia en la investigación cualitativa. En Denman, C. y Haro, J. (comps.), *Por los rincones. Antología de métodos cualitativos en la investigación social*. Hermosillo, Sonora: El Colegio de Sonora, pp. 113-145.
- Guerra, V., Ochoa, S., 2008. Evaluación del programa de manejo de la Reserva de la Biosfera Pantanos de Centla en Tabasco, México. *Univ. Cienc.* 24(2), 135-146.
- Guevara, S., 2010. Las reservas de biosfera en Iberoamérica. *Ambienta*, 92, 46-56.
- Guijt, I., 2008. *Seeking Surprise: Rethinking monitoring for collective learning in rural resource management*. Tesis doctoral publicada, Universidad Wageningen, Holanda.
- 2010. *Accountability and Learning. Exploding the Myth of Incompatibility between Accountability and learning*. En Ubels J., Acquaye N. y Fowler A. (eds.) *Capacity Development in Practice*, Londres y Washington: Earthscan.
- Halachmi A. (2002). Performance Measurement, Accountability, and Improved Performance. *Public Performance & Management Review*, 25(4), 370.
- Hall, A., Sulaiman, V., Clark, N., y Yoganand, B., 2003. From measuring impact to learning institutional lessons: an innovation systems perspective on improving the management of international agricultural research. *Agricultural systems*, 78 (2), 213-241.
- Hansen, M., Alkin, M. y Wallace, T., 2013. Depicting the logic of three evaluation theories. *Evaluation and program planning*, 38, 34-43.
- Harter, S., 1978. Pleasure Derived from Challenge and the Effects of Receiving Grades on Children's Difficulty Level Choices. *Child Development*, 49 (3), 788-799.
- Harvey, C., Saborio, M., Martinez, M., Viguera, B., Chain, A., Vignola, R., y Alpizar, F., 2018. Climate change impacts and adaptation among smallholder farmers in Central America. *Agriculture & Food Security*, 7(1), 57.
- Hatry, H. y Newcomer, K., 2010. Pitfalls in evaluations. En Wholey, J., Hatry, H. y Newcomer, K. (eds.) *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.
- Hernandez, K., Ramalingam, B., Wild, L., 2019. Towards evidence-informed adaptive management. ODI Working Paper 565. ODI, Londres.

- Hicks D., Wouters P., Waltman L., de Riicke S., Rafols I., 2015. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429–431.
- Hockings, M., Stolton, S., Dudley, N., Leverington, F. y Courrau, J., 2006. Evaluating effectiveness: a framework for assessing the management of protected areas. IUCN: Gland, Suiza.
- Hockings, M., Leverington, F. y Cook, C., 2015. Protected area management effectiveness. En Worboys, G., Lockwood, M., Kothari, A., Feary, S. y Pulsford, I. (eds), *Protected Area Governance and Management*, ANU Press, Canberra., pp. 889–928.
- Huitrón, S., 2019. Evaluación de las actividades de restauración ecológica en la reserva de la biosfera maribosa monarca, México 2012-2017. Tesis de licenciatura. Universidad Autónoma del Estado de México.
- Hummelbrunner, R., 2015. Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence. *IDS Bulletin* 46 (1): 17-19.
- InterAcademy Partnership, 2018. Opportunities for Future Research and Innovation on Food and Nutrition Security and Agriculture. The Interacademy Partnership's Global Perspective. Synthesis by IAP Based on Four Regional Academy Network Studies. Trieste; Washington: InterAcademy Partnership.
- Jabareen, Y., 2009. Building a Conceptual Framework: Philosophy, Definitions, and Procedure. *International journal of qualitative methods*, 8 (4), 49-62.
- Jacobson, C., Carter, R., Hockings, M., y Kelman, J., 2011. Maximizing conservation evaluation utilization. *Evaluation*, 17, 1–19.
- Johnson, K., Greenseid, L., Toal, S., King, J., Lawrenz, F. y Volkov, B., 2009. Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. *American Journal of Evaluation*, 30 (3), pp. 377–410.
- Kapos, V., Balmford, A., Aveling, R., Bubb, P., Carey, P., Entwistle, A., Hopkins, J., Mulliken, T., Safford, R., Stattersfield, A., Walpole, M., Manica, A., 2009. Outcomes, not implementation, predict conservation success. *Oryx* 43, 336-342.
- Kessler, A., 2006. Moving people towards collective action in soil and water conservation: experiences from the Bolivian mountain valleys (Tesis de doctorado). Universidad Wageningen, Holanda.
- King, J. 1998. Making sense of participatory evaluation on practice. *New Directions in Evaluation*, 80, 57-67.
- King, J. y Stevahn, L., 2013. *Interactive Evaluation Practice: Mastering the Interpersonal Dynamics of Program Evaluation*. Thousand Oaks: Sage.
- King, J. y Alkin, M., 2019. The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40 (3), 431–458.
- Kirkhart, K., 2000. Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation*, 88, pp. 5–23.
- Korten, D., y Klauss, R. (eds), 1984. *People-Centered Development: Contributions Toward Theory and Planning Frameworks*. West Hartford, Connecticut: Kumarian Press.

- Kumar, S., Coombes, Y., Vovides, Y., y Crabbe, R., 2018. The art of knowledge exchange: a results-focused planning guide for development practitioners in the Social, Urban, Land, and Resilience Sectors. Washington, DC: World Bank.
- Kusters, C., van Vugt, S., Wigboldus, S., Williams, B. y Woodhill, J., 2011. Making evaluations matter: A practical guide for evaluators. Wageningen, Holanda, Centre for Development Innovation, Wageningen University.
- Lamm, A., Israel, G., y Harder, A., 2011. Getting to the bottom line: How using evaluation results to enhance extension programs can lead to greater levels of accountability. *Journal of Agricultural Education*, 52, 44–55.
- Lennie, J., Tacchi, J., Wilmore, M. y Koirala, B., 2015. A holistic, learning-centred approach to building evaluation capacity in development organizations. *Evaluation*, 21 (3), 325-343.
- Lepori, B., P. van den Besselaar, M. Dinges, B. Poti, E. Reale, S. Slipersæter, J. Thèves y B. van der Meulen, 2007. Comparing the Evolution of National Research Policies: What Patterns of Change? *Science and Public Policy*, 34, 372-388.
- Leverington, F., Costa, K., Pavese, H., Lisle, A., y Hockings, M., 2010. A global analysis of protected area management effectiveness. *Environmental Management*, 46(5), 685–698.
- Levine, C., 2007. Catholic Relief Services' (CRS) Guidance for Developing Logical and Results Frameworks. Baltimore: Catholic Relief Services.
- Leviton, L. y Hughes, E., 1981. Research on the Utilization of Evaluations: a Review and Synthesis. *Evaluation Review*, 5 (4), pp. 525–548.
- Lewis, D., 2001. La cooperación como proceso. Construcción de la etnografía institucional en Bangladesh. *Economía, Sociedad y Territorio*, 3(10), pp. 355-377.
- Lewis, D., Bebbington A., Batterbury S., Shah A., Olson E., Siddiqi S. y Duvall S., 2003. Practice, power and meaning: frameworks for studying Organizational culture in Multi-agency rural development projects. *Journal of International Development*, 15, 541-557.
- Lincoln, Y., y Guba, E., 2004. The roots of fourth generation evaluation: Theoretical and methodological origins. En Alkin, M., (Ed.), *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage, pp. 225–242
- Llorente, J., y Ocegueda, S., 2008. Estado del conocimiento de la biota, en capital natural de México. Vol. I: Conocimiento actual de la biodiversidad. México: CONABIO, pp. 283-322.
- LSE Public Policy Group, 2011. *Maximizing the Impacts of Your Research: a Handbook for Social Scientist*. Reino Unido: Public Policy Group.
- Mackay, F., 2005. Draft World Bank Operational Policy 4.10 on Indigenous Peoples: Progress or More of the Same. *Ariz. J. Int'l & Comp.*, 22, p. 65.
- Mackay, R., y Horton, D., 2003. Expanding the use of impact assessment and evaluation in agricultural research and development. *Agricultural systems*, 78 (2), 143-165.
- Maldonado, C. y Galíndez, C., 2013. *Monitoreo, Evaluación y Gestión por Resultados. Aprendizaje y Cooperación Sur-Sur para la Innovación: El papel de los actores subnacionales*, México: CIDE-Centro CLEAR para América Latina.

- Margoluis, R., C. Stem, V. Swaminathan, M. Brown, A. Johnson, G. Placci, N. Salafsky, y I. Tilders, 2013. Results chains: a tool for conservation action design, management, and evaluation. *Ecology and Society* 18(3): 22.
- Mathews, A., 2011. *Instituting Nature: Authority, Expertise, and Power in Mexican Forests*. Cambridge: MIT Press.
- Mayne, J., 2006. Studies are not enough: The necessary transformation of evaluation. *Can. J. Program Eval.* 21, 93-120.
- 2010. Building an evaluative culture: the key to effective evaluation and results management. *The Canadian Journal of Program Evaluation*, 24 (2), 1-30
- McKinnon, M., Mascia, M., Yang, W., Turner, W., y Bonham, C., 2015. Impact evaluation to communicate and improve conservation non-governmental organization performance: the case of Conservation International, *Phil. Trans. R. Soc. B*, 370: 20140282.
- McKinnon, M., Cheng, S., Dupre, S., Edmond, J., Garside, R., Glew, L., y Woodhouse, E., 2016. What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. *Environmental Evidence*, 5(1), 1-25.
- Meier, K. y Krause, G., 2003. *The Scientific Study of Bureaucracy: An Overview*. En Krause, G., y Meier, K. (Eds.). *Politics, policy, and organizations: Frontiers in the scientific study of bureaucracy*. University of Michigan Press.
- Mendizabal, E., 2010. *The Alignment, Interest, and Influence Matrix (AIMM)*. Londres: ODI.
- Merino, L. y Ortiz, G., 2013. *Encuentros y desencuentros. Las comunidades forestales y las políticas públicas en tiempos de transición*. Miguel Ángel Porrúa-Universidad Nacional Autónoma de México, México.
- Mittermeier, R. y Mittermeier, G., 1997. *Megadiversity. Earth's biologically wealthiest nations*. México: CEMEX.
- Mora, C., y Antonie, R., 2012. Levers supporting program evaluation culture and capacity in Romanian public administration: The role of leadership. *Society and Economy*, 34(3), 423–432. doi:10.1556/SocEc.34.2012.3.4
- Mosse, D., 2004. Is Good Policy Unimplementable? Reflections on the Ethnography of Aid Policy and Practice. *Development and Change*, 35 (4): 639–67.
- Mosse, D., 2005. *Cultivating development: an ethnography of aid policy and practice*. Londres: Pluto Press.
- Mosse, D. y Lewis, D., 2006. Theoretical approaches to Brokerage and translation in development. En Lewis, D. y Mosse, D. (eds.) *Development Brokers and Translators: The Ethnography of Aid and Agencies*, Bloomfield: Kumarian Press.
- Mwampamba, T., Maldonado, O., y Richards, M., 2015. *Evaluación participativa de los impactos sociales y de biodiversidad de intervenciones REDD+: Orientación para los facilitadores*. México, D. F: Alianza México de REDD+.
- Nagendra, H., 2018. The Global South Is Rich in Sustainability Lessons That Students Deserve to Hear. *Nature* 557 (7706): 485–488.

- Natsios, A., 2010. The Clash of the Counter-bureaucracy and Development. Center for global development
- Nature, 2020. Ending hunger: Science must stop neglecting smallholder farmers. *Nature*, 586 (7829), 336–336.
- Netherlands Organization for Scientific Research, s.f. Impact and Uptake of Research. Recuperado de <https://www.nwo.nl/en/documents/wotro/impact-and-research-uptake>
- Nowotny, H., Scott, P., y Gibbons, M., 2003. Introduction: 'Mode 2' revisited: The new production of knowledge. *Minerva*, 41 (3), 179-194.
- Nunneley, R., King, J., Johnson, K. y Pejsa, L., 2015. The value of clear thinking about evaluation theory. Example of use and influence. En Christie, C. y Vo, A. (eds.), *Evaluation use and decision making in society. A tribute to Marvin Alkin*. Charlotte, Estados Unidos: Information Age Publishing.
- Nussbaum, B., 2013. Stop Fetishizing Failure. *Creative Intelligence*. Recuperado de: <https://creativeintelligencebook.com/post/41570210573/stop-fetishizing-failure>
- Okail, N., 2009. *Dancing with the Experts An Activist Research for Legitimacy in Egypt's Aid Relations*. Tesis doctoral. Universidad de Sussex, Inglaterra.
- Olsen, W., 2014, Agosto 27. What is Qualitative Comparative Analysis? [Video]. Youtube. <https://www.youtube.com/watch?v=q917kpR9iGM&t=235s>
- Olvera, J., Piña, H., y Mercado, A., 2009. La universidad pública: autonomía y democracia. *Convergencia Revista de Ciencias Sociales*, 51, 237-254.
- Organización Latinoamericana y del Caribe de Entidades Fiscalizadoras Superiores, 2015. *Áreas Protegidas América Latina. Auditoría coordinada*. Brasilia, Brasil: Organización Latinoamericana y del Caribe de Entidades Fiscalizadoras Superiores, Comisión Técnica de Medio Ambiente (COMTEMA), Tribunal de Cuentas de la Unión, 32 pp.
- Orta, M., Cardozo, M., Méndez, H., Arredondo, C. y Espejel, I., 2013. Análisis del sistema de evaluación de un programa ambiental de la política mexicana: el Proders y su transformación al Procodes. *Investigación ambiental. Ciencia y Política Pública*, 5(1), 44-61.
- Oslo Institute for Research on the Impact of Science, 2016. Introduction to the new centre OSIRIS – The OSlo Institute for Research on the Impact of Science. Recuperado de: <https://www.sv.uio.no/tik/english/research/centre/osiris/presentations/presentations-2017/osiris-presentation-general-english.pdf>
- Owen, H., 2008. *Open space technology: A user's guide*. Berrett-Koehler Publishers.
- Pasanen, T., Barnett, I. y Buffardi, A. (2019). *Supporting adaptive management: monitoring and evaluation tools and approaches*. Londres: GLAM, Overseas Development Institute
- Patton, M., 1997. *Utilization-focused evaluation: The new century text*. Thousand Oaks: Sage.
- 2010. *Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use*. Nueva York: Guilford Press.
- Patton M. y Horton, D., 2009. *Utilization-focused evaluation for agricultural innovation*, ILAC Brief 22.

Roma: ILAC, Bioversity.

- Peersman, G., Guijt, I. y Pasanen, T., 2015. Evaluability assessment for impact evaluation. Guidance, checklist and decision support for those conducting the assessment. A Methods Lab publication, Londres: Overseas Development Institute.
- Pérez, G. (2015, noviembre), Cuando la evaluación funciona (o no) como motor de cambio en las políticas públicas: el diseño institucional de la evaluación y los efectos en su utilización, ponencia presentada en el XX Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Lima.
- Perreault, T., 2009. Environment and Development. En Castree, N., Demeritt, D., Liverman, D. y Rhoads, B. (Eds.), A companion to environmental geography. Wiley-Blackwell, Singapur. pp. 442-46.
- Perrin, B., 2002. How to – and How Not to – Evaluate Innovation, *Evaluation*, 8 (1), pp. 13–28.
- Pigliucci, M., 2017. The Virtuous Skeptic. *Skeptical Inquirer*, 41 (2): 54–57.
- Plaza, J., 2020. Análisis de propuestas para fortalecer el presupuesto de la CONANP [Diapositivas de PowerPoint]. Recuperado de http://comisiones.senado.gob.mx/ambiente/reu/docs/presentacion1_e100220.pdf
- Poister, T. y Streib, G., 1999. Performance Measurement in Municipal Government: Assessing the State of the Practice. *Public Administration Review* 59 (4): 325 – 35.
- Polastri, R. y Savastano, S., 2020. Financiación basada en los resultados: Un posible cambio de juego para las futuras operaciones del FIDA. Recuperado de: <https://www.ifad.org/es/web/latest/blog/asset/42102349>
- Pomeroy, R., Parks, J., y Watson, L., 2004. How is your MPA doing? A guidebook of natural and social indicators for evaluating marine protected area management effectiveness, IUCN.
- Preskill, H. y Boyle, S., 2008. A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation*, 29 (4), 443–459.
- Preskill, H. y Torres, R., 1999. Building capacity for organizational learning through evaluative inquiry. *Evaluation*, 5, 42–60.
- Preskill, H. y Torres, R., 2000. The learning dimension of Evaluation Use. *New directions for evaluation*, 88, pp. 25–37.
- Pressey, R., Visconti, P., y Ferraro, P., 2015. Making parks make a difference: Poor alignment of policy, planning and management with protected-area impact, and ways forward. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1681), 20140280.
- Prieto M., Hernandez, K., Faith, B. y Ramalingam, B., 2017. Doing digital development differently: lessons in adaptive management from technology for governance initiatives in Kenya. Making All Voices Count Research Report. Brighton: Institute of Development Studies.
- Raeside, A., 2011. Are INGOs brave enough to become learning organisations? En Ashley, H., Kenton, N. y Milligan, A. (eds.), *How Wide are the Ripples? From Local Participation to International Organisational Learning*. Participatory Learning and Action, 63: 97-102.

- Rafols, I. (26 de octubre de 2017). There's no silver bullet for measuring societal impact. [Mensaje en un blog]. OSIRIS blog. Recuperado de: <https://www.sv.uio.no/tik/english/research/centre/osiris/osirisblog/there%E2%80%99s-no-silver-bullet-for-measuring-societal-im.html>
- Rafols, I., Robinson, N. y van Leeuwen T., (23 de marzo de 2017). How to make altmetrics useful in societal impact assessments: shifting from citation to interaction approaches. [Mensaje en un blog]. Impact of Social Sciences. Recuperado de <http://blogs.lse.ac.uk/impactofsocialsciences/2017/03/23/how-to-make-altmetrics-useful-in-societal-impact-assessments-shifting-from-citation-to-interaction-approaches/>.
- Ramírez, H., Pedroza, A., Martínez, J., y Valdez, R., 2011. Participatory assessment of the soil degradation in the mapimí biosphere reserve. *Revista Chapingo Serie Horticultura*, XVII.
- Ramos, D., Álvarez, M., Morales, M., 2018. Sustentabilidad y patrimonio biocultural en la Reserva de la biosfera El Ocote. *Trace* 74, 9-37.
- Ramos I., D'Este P., Woolley R. y Amara N. (2018). Introduction to a special section: Balancing scientific and societal impact—A challenging agenda for academic research. *Science and Public Policy*, scy033, <https://doi.org/10.1093/scipol/scy033>
- Reeler, D., 2007. *A Theory of Social Change and Implications for Practice, Planning, Monitoring and Evaluation*. Community Development Resource Association.
- Reddy, V. y Wilson, C., 2015, Junio 8. Session 10, Reddy & Wilson - Humility and Openness to Engagement. [Video]. Youtube. <https://www.youtube.com/watch?v=3imloUK-DYE>
- Richards, M., 2012. La evaluación participativa del impacto social para proyectos y programas de recursos naturales, *Forest Trends*. Recuperado de http://www.forest-trends.org/documents/files/doc_3828.pdf.
- Richards, M. y Panfil, S., 2011. *Manual Para la Evaluación de Impacto Social y sobre la Biodiversidad (EISB) para Proyectos REDD+: Parte 1 – Guía Básica para los Proponentes de Proyectos*. Washington, DC: Alianza para el Clima, Comunidad y Biodiversidad, Forest Trends, Fauna & Flora International y Rainforest Alliance.
- Rivera, R., Dutrénit, G., Ekboir, J., Vera-Cruz, A., 2016. Vinculación y productividad de la investigación: una comparación de diferentes perfiles de investigadores. En Vera-Cruz, A. y Dutrénit, G. (eds.). *Sistema de innovación del sector agropecuario en México. Tendiendo puentes entre los actores*. Ciudad de México, Miguel Ángel Porrúa-Universidad Autónoma Metropolitana, 143-172.
- Rocha, A. y De la Cerda, F., 2016. Principales actores del Sistema de Innovación Agropecuario Mexicano. En Vera-Cruz, A. y Dutrénit, G. (eds.). *Sistema de innovación del sector agropecuario en México. Tendiendo puentes entre los actores*. Ciudad de México, Miguel Ángel Porrúa-Universidad Autónoma Metropolitana, 73-104.
- Roe, D., Grieg, M. y Mohammed, E., 2013. *Assessing the social impacts of conservation policies: rigour versus practicality*. IIED Briefing Papers. International Institute for Environment and Development.
- Roetman, E., 2011. *A can of worms? Implications of rigorous impact evaluations for development agencies*. Working paper 11. Nueva Delhi, International Initiative for Impact Evaluation.
- Rousseau, I., 2017. *La nueva regulación de la gestión social de los proyectos energéticos en México*.

Seguridad, sustentabilidad y gobernabilidad. *Revista mexicana de ciencias políticas y sociales*, 62(230), 197-219.

Salafsky, N., Margoluis R., y Redford K. (2001). *Adaptive management: a tool for conservation practitioners*. Washington, D.C, Estados Unidos: Biodiversity Support Program

Salafsky, N., Salzer, D., Stattersfield, A., Hilton, C., Neugarten, R., Butchart, S., Collen, B., Cox, N., Master, L., O'Connor, S. y Wilkie, D., 2008. A standard lexicon for biodiversity conservation: unified classifications of threats and actions. *Conservation Biology* 22:897-911.

Salafsky, N., Boshoven, J., Burivalova, Z., Dubois, N., Gomez, A., Johnson, A. y Wordley, C., 2019. Defining and using evidence in conservation practice. *Conservation Science and Practice*, 1(5), e27.

Salzer, D., y Salafsky, N., 2006. Allocating resources between taking action, assessing status, and measuring effectiveness of conservation actions. *Nat. Areas J.* 26, 310-316.

Sánchez, V., Figueroa, F., Illoldi, P. y Linaje, M., 2011. Efectividad del sistema de áreas protegidas para conservar la vegetación natural. En Koleff, P. y Urquiza, T. (coords.) *Planeación para la conservación de la biodiversidad terrestre en México: retos en un país megadiverso*. DF, Mexico: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad–Comisión Nacional de Áreas Naturales Protegidas, pp. 59–86

Sanderson, I., 2000. Evaluation in Complex Policy Systems, *Evaluation*, 6 (4), pp. 433–454.

Sarewitz, D., 2016. Saving science. *The New Atlantis*, 49, 4-40.

Schön, D., 1995. Knowing-in-action: The new scholarship requires a new epistemology. *Change: The Magazine of Higher Learning*, 27 (6), 27-34.

Science Europe, 2020. Recommendations on research assessment processes. Recuperado de: <https://www.scienceeurope.org/media/3twjxim0/se-position-statement-research-assessment-processes.pdf>

Scott, J., 1998. *Seeing like a state. How certain schemes to improve the human condition have failed*. Yale University Press.

Scriven, M., 1967. *The Methodology of Evaluation*. En Tyler, R., Gagne, R. y Scriven, M. (eds.), *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally & Co.

----- 2005. Review of the book: *Empowerment Evaluation Principles in Practice*. *American Journal of Evaluation*, 26 (3), pp. 415–417.

----- 2007. *The logic of evaluation*. En Hansen, H. (ed.), *Dissensus and the Search for Common Ground*. Windsor: OSSA. PP. 1-16

Secretaría de Educación Pública (2018). Reglas de operación del programa para el desarrollo profesional docente para el ejercicio fiscal 2018. En *Diario Oficial de la Federación*, 27 de diciembre de 2017, México. Recuperado de http://prodep.estrategianacionaldeformaciondocente.sems.gob.mx/regop/Reglas_2018.pdf

Segerholm, C., 2002. Evaluation as responsibility, conscience, and conviction. En Ryan, K., y Schwandt, T., (Eds.), *Exploring evaluator role and identity* Greenwich, CT: Information Age Publishing Inc., pp. 87–102.

SEMARNAT, s.f. Impacto ambiental. Consulta Temática. Recuperado de: http://dgeiawf.semarnat.gob.mx:8080/ibi_apps/WFSerMet?IBIF_ex=D4_R_IMPACTO00_01&IBIC_user=dgeia_mce&IBIC_pass=dgeia_mce

Serrat, O., 2010. Knowledge Solutions. Asian Development Bank, Washington D.C.

Shukla, P., Skea J., Calvo E. *et al.*, 2019. IPCC, 2019. climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Intergovernmental Panel on Climate Change (IPCC).

Shulha, L. y Cousins, B. 1997. Evaluation use: Theory, research, and practice since 1986. *American Journal of Evaluation*, 18, pp. 195–208.

Shulha L., Whitmore, E., Cousins, B., Gilbert, N. y Hudib, H., 2016. Introducing Evidence-Based Principles to Guide Collaborative Approaches to Evaluation: Results of an Empirical Process. *American Journal of Evaluation*, 37 (2), 193-215.

Silva, C. y Castro, A., 2014. Los cuerpos académicos, recursos y colegialidad forzada. *Archivos Analíticos de Políticas Educativas*, 22 (68), 1-26.

Snowden, D. y Boone, M., 2007. A Leader's Framework for Decision Making. *Harvard Business Review*, pp. 69-76.

Sotelo, E., Rosas, A. y Rodríguez, L., 2015. Capítulo XI. Medio Ambiente. En Cardozo, M. (coord.), *Institucionalización de procesos de evaluación. Calidad y utilización de sus resultados*. Ciudad de México: Universidad Autónoma Metropolitana, pp. 143-158.

Spaapen, J., Van Drooge L., Propp T., van der Meulen B., Shinn T., Marcovich A., van den Besselaar P., de Jong S., Barker K., Cox D., y Morrison K., 2011. *Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society*. Recuperado de http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf

Spaapen, J., y van Drooge, L., 2011. Introducing “productive interactions” in social impact assessment. *Research Evaluation*, 20, 211–218.

Stairs, G., 2007. *The social construction of sustainable rural futures in the world network of biosphere reserves*. Tesis doctoral, University of Wisconsin–Madison.

Stake, R., 1986. *Quieting reform: Social science and social action in an urban youth program*. Urbana: University of Illinois Press.

Stem, C., R. Margoluis, N. Salafsky, y M. Brown, 2005. Monitoring and evaluation in conservation: a review of trends and approaches. *Conservation Biology*, 19 (2): 295-309.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., y Befani, B., 2012. *DFID Working Paper 38. Broadening the range of designs and methods for impact evaluations*. Londres: DFID

Stewart, J., 2014. *Developing a culture of evaluation and research*. Paper N. 28. Child Family Community Australia.

Steyaert, P. y Jiggins, J., 2007. Governance of complex environmental situations through social learning: a synthesis of SLIM's lessons for research, policy and practice. *Environmental*

Science and Policy, 10(6), pp. 575-586.

Stockdill, S., Baizerman, M., y Compton, D., 2002. Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7-25.

Stoll, S., 2010. Evaluation of management effectiveness in protected areas: Methodologies and results. *Basic and Applied Ecology*, 11(5), 377–382.

Stolton, S. y Dudley, N., 2016. *METT Handbook: A guide to using the Management Effectiveness Tracking Tool (METT)*. Woking, Inglaterra: WWF-UK, 74 pp.

Strathern M., 1997. 'Improving ratings': audit in the British University system. *European Review*, 5, 305-321.

Suárez, M. y Muñoz, H., 2004. Ruptura de la institucionalidad universitaria. En Ordorika I. (Coord.). *La academia en jaque: Perspectivas políticas sobre la evaluación de la educación superior en México*. México: Universidad Nacional Autónoma de México, Miguel Angel Porrúa, 25–33

Taut, S., 2007. Studying self-evaluation capacity building in a large international development organization, *American Journal of Evaluation*, 28, 45–59.

Taylor, T., Suarez, Y., Garcia, E., Henry, D. y Balcazar, F., 2013. Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation*, 34 (2), 190-206.

Tilley H, Ball L. y Cassidy C., 2018. *Research Excellence Framework (REF) Impact Toolkit*. Overseas Development Institute.

Torres, R. y Preskill, H., 2001. Evaluation and organizational learning: Past, Present and Future, *American Journal of Evaluation*, 22 (3), pp. 387–395.

Turnhout, E., 2009. The effectiveness of boundary objects: The case of ecological indicators. *Science and Public Policy*, 36(5), 403–412. <https://doi.org/10.3152/030234209X442007>

Turnhout E., Stuver, M., Klostermann, J., Harms, B., y Leeuwis, C., 2013. New roles of science in society: Different repertoires of knowledge brokering. *Science and Public Policy*, 40, pp. 354–365.

Turnhout, E., Skutsch, M., y de Koninck, J., 2015. Carbon accounting. En Bäckstrand, K. y Löwbrand, E. (eds.) *Research Handbook on Climate Governance*. Cheltenham: Edward Elgar Publishing.

Turnhout E. (2018). The Politics of Environmental Knowledge. *Conservation and Society*, 16 (3), 363-371

Turnhout, E., McElwee, P., Chiroleu-Assouline, M., Clapp, J., Isenhour, C., Kelemen, E., Jackson, T., Miller, D. C., Rusch, G. M., Spangenberg, J. H., & Waldron, A., 2021. Enabling transformative economic change in the post-2020 biodiversity agenda. *Conservation Letters*, 14(4). <https://doi.org/10.1111/conl.12805>

Turri, J., Alfano, M. y Greco, J., 2019. Virtue Epistemology. En Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*. Recuperado de <https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/>.

- UNDP, 2020. The next frontier-Human development and the Anthropocene. Human Development Report 2020.
- UNESCO, 2013. The world network of biosphere reserves. Ecological Sciences for Sustainable Development. Recuperado de: <http://www.unesco.org/new/en/natural-sciences/environment/ecological-sciences/man-and-biosphere-programme/>
- UNESCO, 2019. Dieciocho nuevas reservas ingresan en la Red Mundial de Reservas de Biosfera de la UNESCO. Recuperado de: <https://es.unesco.org/news/dieciocho-nuevas-reservas-ingresan-red-mundial-reservas-biosfera-unesco>
- Urquijo, P., 2014. El paisaje como concepto geográfico, histórico y ambiental. En Barrera, S. y Monroy, J. (eds.). *Perspectivas sobre el paisaje*. Bogotá: Universidad Nacional de Colombia.
- Van der Meulen, B., Van Drooge L., van den Besselaar P., Elsen M., de Haas M., van den Heuvel J. y Maassen H., 2010. Evaluating the societal relevance of academic research: a guide. Recuperado de <https://repository.tudelft.nl/islandora/object/uuid:8fa07276-cf52-41f3-aa70-a71678234424>
- Van Hemelrijck, A. y Guijt, I., 2016. *Balancing Inclusiveness, Rigour and Feasibility: Insights from Participatory Impact Evaluations in Ghana and Vietnam*. CDI Practice Paper 14, Brighton: IDS
- Vallaey, F., 2014. La responsabilidad social universitaria: un nuevo modelo universitario contra la mercantilización. *Revista Iberoamericana de Educación Superior*, 5 (12), 105-117.
- Vanclay, F., 2003. *International Principles For Social Impact Assessment*. *Impact Assessment and Project Appraisal*, 21 (1), 5-12.
- Vasen, F. y Lujano I., 2017. Sistemas nacionales de clasificación de revistas científicas en América Latina: tendencias recientes e implicaciones para la evaluación académica en ciencias sociales. *Revista Mexicana de Ciencias Políticas y Sociales*, 231, 199-228.
- Vázquez, C., y Fermán, J., 2010. Evaluación del impacto socioeconómico de la Reserva de la Biosfera Alto Golfo de California y Delta del Río Colorado en la actividad pesquera ribereña de San Felipe, Baja California, México. *Región y sociedad*, 22 (47), 31-51.
- Vedung, E., 2010. Four Waves of Evaluation Diffusion, *Evaluation*, 16: 263.
- Vera, A., Dutrénit, G., Ekboir, J., Martínez, G. y Torres, A., 2008. Virtues and Limits of Competitive Funds to Finance Research and Innovation: The Case of Mexican Agriculture. *Science and Public Policy*, 35, 501-513.
- Vera, A., y Dutrénit, G., 2016. Tendiendo puentes entre los actores del sistema de innovación del sector agropecuario: una introducción. En Vera-Cruz, A. y Dutrénit, G. (eds.). *Sistema de innovación del sector agropecuario en México. Tendiendo puentes entre los actores*. Ciudad de México, Miguel Ángel Porrúa Universidad Autónoma Metropolitana.
- Vermeulen, S., Woodhill, J., Proctor, F.J. y Delnoye, R., 2008. *Chain-wide learning for inclusive agrifood market development: A guide to multi-stakeholder processes for linking small-scale producers with modern markets*. Reino Unido: International Institute for Environment and Development (IIED).
- Vidaurri, S. y Luna B., 2017. Del Programa Integral de Fortalecimiento Institucional (PIFI) al Programa de Fortalecimiento de la Calidad Educativa (PFCE), su efecto en la universidad pública

- estatal. *Revista Dilemas Contemporáneos: Educación, Política y Valores*, 5 (1), 1-31.
- Visser I., Guijt, I., Kusters, C., Roefs, M., de Ruyter, M. y Wolsky, A., 2013. *Impact Evaluation: Taking stock and looking ahead*. Wageningen, Holanda: Centre for Development Innovation.
- Vo, A., 2015. Foreword. En Christie, C. y Vo, A. (eds.), *Evaluation use and decision making in society. A tribute to Marvin Alkin*. Charlotte, Estados Unidos: Information Age Publishing.
- Walford, G., 2005. Research ethical guidelines and anonymity. *International Journal of Research & Method in Education*, 28(1), 83–93. <https://doi.org/10.1080/01406720500036786>
- Weißhuhn, P., Helming, K., y Ferretti, J., 2018. Research impact assessment in agriculture—A review of approaches and impact areas. *Research Evaluation*, 27(1), 36-42.
- Weiss, C., 1977. *Using social research in public policy making*. Heath, Estados Unidos: Lexington Books.
- 1982. Policy Research in the Context of Diffuse Decision Making. *The Journal of Higher Education*, 53 (6), 619–639.
- 1997. Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, (76), 41–55.
- 1998. Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, Vol. 19, No. 1, pp. 21-33.
- 1999. The Interface between Evaluation and Public Policy, *Evaluation*, 5 (4): 468-486.
- Weiss, C., Murphy, E. y Birkeland, S., 2005. An alternative route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation*, 26, pp. 12–30.
- Westhorp, G., 2014. *Realist impact evaluation an introduction*. Londres: Overseas Development Institute, pp. 12
- Whitty, B., 2015. Mapping the results landscape: Insights from a crowdsourcing survey. En Eyben, R., Roche, C., Guijt, I. y Shutt, C. (eds.), *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.57-78. Rugby, UK: Practical Action Ltd.
- Wild, L. y Ramalingam, R., 2018. *Building a global learning alliance on adaptive management*. ODI, Londres. Recuperado de: www.odi.org/publications/11167-building-global-learning-alliance-adaptive-management
- Williams, B., 2015. Prosaic or Profound? The Adoption of Systems Ideas by Impact Evaluation. *IDS Bulletin* 46 (1): pp- 7-16
- Willis, P., y Trondman, M., 2000. Manifesto for ethnography. *Ethnography*, 1(1), 5–16.
- Woodhouse, E., Homewood, K., Beauchamp, E., Clements, T., McCabe, J., Wilkie, D., y Milner, E., 2015. Guiding principles for evaluating the impacts of conservation interventions on human well-being. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370 (1681), 20150103.
- Woodhouse, E., Bedelian, C., Dawson, N., y Barnes, P., 2018. Social impacts of protected areas: exploring evidence of trade-offs and synergies.

En Schreckenberg, K., Mace, G. y Poudyal, M. (Eds.) Ecosystem services and poverty reduction: trade-offs and governance. Routledge: London and New York.

Yarbrough, D., Shulha, L., Hopson, R. y Caruthers, F., 2011. The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users. Estados Unidos: Sage Publications.

Yin, R., 2009. Case Study Research: Design and Methods. Thousand Oaks, Estados Unidos: Sage Publication.

Anexos

Anexo 1. Guión de entrevista semi-estructurada

Preguntas rompe-hielo

1. ¿Cuándo piensa en la evaluación de proyectos, que palabra le viene a la mente?, ¿Por qué?
2. ¿Cuál fue la mejor evaluación de un proyecto en la que ha participado? ¿Qué hizo que fuera tan buena?, ¿Qué pasó como resultado?

Características de las evaluaciones

3. ¿Es esta una organización donde las evaluaciones son parte rutinaria del trabajo?
4. ¿Cómo se hacen evaluaciones en la organización? ¿Existen procedimientos estandarizados para hacer evaluaciones? ¿Qué metodología utilizan?
5. ¿Para qué se quiere la evaluación?
6. ¿Se ha generado en el pasado información relevante (p. ej. la evaluación permite saber si los proyectos están funcionando, las personas de la organización usan información producto de las evaluaciones como parte de sus actividades cotidianas) a través de evaluaciones?
7. ¿Quién es el usuario principal de las evaluaciones? (Directivo/Donante/Partes interesadas/Personal de la organización)

Tiempo y presupuesto

8. ¿El personal tiene tiempo para hacer actividades de evaluación?
9. ¿Hay presupuesto para las actividades de evaluación (diseño, recolección, análisis, reporte)?
10. Si los resultados de las evaluaciones proponen cambios ¿es posible financiarlos?

Capacidad técnica

11. ¿Qué experiencia tiene en la evaluación de proyectos?
12. ¿Ha recibido capacitación, formal o informal, en torno de cómo hacer y usar evaluaciones?
13. ¿Existe personal dedicado a diseñar, implementar y analizar los resultados de evaluaciones?

Sistemas de soporte – confianza y comunicación

14. ¿A qué grado y de qué manera (formal o informal) la información es compartida al interior de la organización?
15. ¿A qué grado las personas de la organización usan información producto de evaluaciones como parte de sus actividades cotidianas?
16. ¿El entorno de trabajo y las actividades rutinarias fomentan que se hable y analice lo que está pasando con los proyectos? ¿Qué obstáculos existen para permitir la frecuente y sencilla comunicación con los demás?
17. ¿Se le da al personal el tiempo suficiente y los espacios para reflexionar en torno de los éxitos y fracasos de la organización?
18. ¿Quién trabaja en esta organización confía en que errores o fallas no los afectarán negativamente?

Rendición de cuentas

19. ¿Las evaluaciones son respuesta de un mandato o de un financiamiento recibido? ¿Esta solicitud determina las características de la evaluación? (tiempos, formato, metodología, información por recolectar)
20. ¿A quién se le rinde cuentas? ¿Para que las quieren los que las solicitan? ¿Qué información solicitan?
21. ¿Las expectativas de aquellos a los que se les rinde cuentas afectan lo que se decide reportar a través de las evaluaciones?
22. ¿Cómo reaccionaría/ha reaccionado aquel que pide que se le rindan cuentas ante resultados no previstos? ¿Existe presión por lograr impactos predeterminados?

Diseño de proyectos⁸³

23. ¿Está claramente definida la idea de cómo lo que hará el proyecto conducirá a lo que se quiere conseguir?
24. ¿Está claramente definida la teoría del cambio del proyecto?
25. ¿El proyecto cuenta con indicadores para establecer *outputs* y *outcomes*?
26. ¿El proyecto contempla una planeación de actividades de monitoreo y evaluación?

⁸³ Información obtenida principalmente a través de la revisión de la documentación de los proyectos.

Anexo 2. Guion para el acercamiento a las organizaciones

Busco mejorar y adaptar los métodos de evaluación que usamos a través de colaboraciones entre académicos y practicantes. En ese sentido, me gustaría ofrecer capacitaciones que les permitan cumplir con sus propios objetivos. Para identificar las áreas en las que podría colaborar, la idea sería entender primero si están satisfechos con la forma en que practican sus evaluaciones, a qué tipo de limitaciones se enfrentan y en qué creen que podrían mejorar.

Un ejemplo de lo que podríamos hacer es que les acompañe en el proceso de hacer la evaluación de uno de sus proyectos. En ese contexto podrían darse asesorías y organizar sesiones de trabajo. Esto permitiría enseñar varias cosas sobre el proceso de evaluación a través de un caso práctico que se combine bien con su agenda de trabajo.

Le aseguro que, a menos de que usted lo consienta, en los productos académicos que surjan como resultado de esta colaboración resguardaré información para garantizar el anonimato de la organización.

En resumen, mi objetivo no es evaluar sus proyectos, sino colaborar con ustedes para saber cómo podemos mejorar la forma en que nos damos cuenta de si lo que estamos haciendo está funcionando. La idea es reunir ese aprendizaje para que en la región diseñemos proyectos que cada vez cumplan mejor con sus objetivos. Me gustaría mucho interactuar con su organización durante este año si cree que eso les sería a ustedes de utilidad.

Anexo 3. Resultados del foro “Evaluación de la efectividad en el manejo de la Reserva”

El foro promovió la discusión de ventajas, retos y posibles fallas de alianzas para evaluar la efectividad en el manejo de la reserva entre integrantes de diferentes organizaciones que manejan proyectos de conservación y desarrollo.

En un primer momento, usando la técnica del café mundial, los participantes conversaron sobre las fortalezas y retos de sus organizaciones en relación con la práctica evaluativa (Imágenes 1 y 2).

Las preguntas que guiaron las rondas de conversación fueron:

- ¿Cómo nos percatamos en la organización de que nuestras acciones están generando un impacto?
- ¿Qué obstáculos existen para que podamos hacer mejores evaluaciones?
- Con la finalidad de aprender juntos ¿cómo podríamos superar estos desafíos y fortalecer una estrategia conjunta de evaluación? ¿Cuál sería el primer paso concreto que debería de realizarse?

Imagen 1. Un equipo compartiendo sus reflexiones durante la actividad del café mundial



Crédito de la fotografía: Eduardo Luna Sánchez

En general, se indicó que el objetivo principal de las evaluaciones es informar que las actividades prometidas se realizan con el dinero asignado para ello. Por tanto, no siguen ninguna metodología para establecer la causalidad. Las evaluaciones se centran en los productos entregables y en su mayoría se realizan para rendir cuentas.

Los participantes coincidieron en que las evaluaciones tienen el potencial de generar conocimiento para mejorar sus proyectos. En ese sentido, el grupo reflexionó sobre los primeros pasos que se deben dar para fortalecer una estrategia de evaluación conjunta en la reserva. La idea impulsora era que las evaluaciones debían servir para compartir las lecciones aprendidas, de forma que las organizaciones pudieran trabajar de manera más eficaz en el futuro. Estos son algunos de los pasos clave que se discutieron:

- Manejo de un idioma común

Se consideró fundamental integrar un lenguaje común en el diseño de proyectos para fortalecer la comunicación, la transversalidad y la sinergia. Tener una visión compartida sobre objetivos, problemas a abordar e indicadores a utilizar facilita el intercambio de información y evita la duplicación de esfuerzos.

- Fortalecimiento del papel del consejo asesor de la reserva

El consejo asesor es un mecanismo organizativo de la CONANP que podría retomarse para:

- Liderar acciones que faciliten el intercambio de información entre organizaciones.
- Liderar la iniciativa de estandarizar elementos del sistema de evaluación de cada organización, para que se maneje un lenguaje común.
- Difundir la información que cada organización recopila al realizar evaluaciones
- Ser un lugar de encuentro para compartir lecciones aprendidas, causas de éxito y causas de fracasos.

- Facilitar la circulación del conocimiento

Para difundir el conocimiento adquirido a través de las evaluaciones, es necesario adaptarlo a diferentes públicos, preferentemente utilizando un lenguaje sencillo que facilite su uso. La información debe responder a las necesidades de quien la utilizará. Con base en esto, se debe elegir el idioma (técnico o común) y los medios de comunicación (centros de información municipal, redes sociales, radio, periódicos, delegados, miembros del consejo asesor, académicos o una página web).

Algunas propuestas específicas fueron:

- I. Crear una plataforma online que almacene información y enlace proyectos, preguntas y personas.
- II. Utilizar espacios formales e informales para tener una comunicación efectiva. Es decir, promover el contacto interpersonal a través de reuniones periódicas (como foros o reuniones de intercambio de experiencias), para que el aprendizaje se transmita a través del diálogo.
- III. Permitir el libre flujo de información entre organizaciones, sin bloquearlo con barreras institucionales.

- Mejora de los mecanismos de evaluación

Se dijo que es necesario adecuar los mecanismos de evaluación de todas las organizaciones involucradas en la gestión de la reserva, para que:

- Las evaluaciones dejen constancia de logros, errores y lecciones aprendidas.
- Se eliminen los sistemas de evaluación punitiva, para que el miedo a fallar no cierre oportunidades de mejora.
- Se establezcan procesos permanentes de evaluación, en lugar de evaluaciones *ex post* desvinculadas de la toma de decisiones.

Imagen 2. Un equipo conversando durante la actividad del café mundial



Crédito de la fotografía: Eduardo Luna Sánchez

Durante la segunda actividad del foro, los participantes conversaron sobre indicadores biofísicos, sociales e institucionales que se podrían incorporar a un sistema de evaluación que oriente las decisiones de manejo en la reserva. Para esto, como paso previo, los participantes trabajaron en mapas conceptuales para identificar relaciones entre objetivos de conservación o desarrollo, amenazas y estrategias.

La idea era asociar un indicador a cada elemento del mapa conceptual y discutir:

I. Si el indicador elegido midió, o no, el grado en que las acciones contribuyeron a lograr un cambio.

II. El costo del monitoreo de ese indicador, considerando los recursos disponibles en nuestras organizaciones.

En la tercera y última actividad del foro, utilizando la técnica del espacio abierto los participantes compartieron ideas sobre temas problemáticos que fueron propuestos por ellos.

Los temas discutidos fueron:

- ¿Cómo están trabajando los consejos municipales de desarrollo rural sostenible y el consejo asesor de la reserva?
- ¿Cómo podemos realizar la consulta pública del nuevo programa de manejo, para que las organizaciones y comunidades lo incorporen en su toma de decisiones?
- ¿Cómo incorporar la agenda de género en lo que hemos discutido?
- ¿Cómo promover una participación genuina de los habitantes de la reserva?

Se diseñó una página web para compartir las ideas discutidas durante el foro (Figura 11). Asimismo, como resultado del encuentro, se creó un grupo de google para facilitar la colaboración de los interesados en la revisión y mejora del borrador del nuevo programa de manejo de la reserva.

Figura 11. Capturas de pantalla de la página web del foro.



Anexo 4. Propuesta de reforma al sistema de evaluación científica de la Universidad

Septiembre de 2020

El documento que se presenta se preparó a solicitud del rector como una propuesta de continuidad para el periodo 2021-2024.

RESUMEN DE LA PROPUESTA

El modelo de gestión organizacional de la Universidad es la Responsabilidad Social Universitaria, por ello, para la Universidad es crucial que la investigación científica se oriente hacia la resolución de problemas sociales. En este contexto, existe un área de mejora en sus procedimientos de evaluación científica, basados principalmente en métricas de productividad académica. Esta propuesta de continuidad presenta una serie de actividades para transformar la cultura de investigación de la Universidad, de forma que rinda cuentas a través de la obtención de resultados de desarrollo. Se espera, para el año 2023, posicionar a la Universidad como un líder internacional en torno de las iniciativas que están impulsando un modelo de ciencia socialmente relevante.

ANTECEDENTES

- I. Para la Universidad, la investigación científica es una actividad que refleja el compromiso que mantiene con su entorno. Esto es así, pues el conocimiento que produce se vincula con los problemas del país. Por ello, uno de los objetivos estratégicos que se ha trazado como organización es lograr la excelencia en investigación con proyección social, de forma que se generen conocimientos para promover el desarrollo económico y el bienestar social a nivel estatal y nacional.
- II. De forma adicional, la Universidad ha impulsado una política para revalorizar la vinculación como función sustantiva de la organización. La Universidad permite el registro y financiamiento de proyectos de vinculación, los cuales tienen el mismo valor que los proyectos de investigación en los tabuladores de evaluación académica. Esto implica su reconocimiento como carga horaria, y su incorporación en convocatorias de obtención de plaza, promoción docente y estímulos. En ese sentido, el procedimiento facilita que los investigadores promuevan el uso del conocimiento para atender "...problemas prácticos y actuales a escala local, regional o nacional en vinculación con los sectores productivo, gubernamental, educativo y la sociedad en general."
- III. Después de todo, las universidades reciben recursos públicos porque se espera que el conocimiento genere efectos positivos para la sociedad (Ramos *et al.*, 2018). Por ello, la universidad pública requiere un modelo de gestión que le permita informar a los contribuyentes sobre la forma en que retribuye la confianza depositada en ella. Siendo consecuente, la administración actual eligió la Responsabilidad Social

Universitaria (RSU) como su modelo de gestión organizacional. Este enfoque ético resalta la importancia de identificar los impactos sociales atribuibles a la universidad (Bacigalupo, 2008; Vallaeys, 2014a).

IV. En este contexto, una piedra angular es la realización de evaluaciones para rendir cuentas y mejorar los proyectos universitarios. Actualmente, ya no se considera suficiente que la evaluación se restrinja a la valoración entre pares o a la presentación de productos académicos. En cambio, cada vez más se espera que las universidades adopten un enfoque orientado a la obtención de resultados de desarrollo, es decir, una ciencia de calidad con pertinencia social (Serrat, 2010; Nowotny *et al.*, 2003; Sarewitz, 2016; Ekboir *et al.*, 2017; CLACSO, 2020a).

PROBLEMA QUE JUSTIFICA LA PROPUESTA DE REFORMA

- V. Dada esa expectativa, la actual administración encuentra una importante área de mejora en los procedimientos de evaluación que utiliza la universidad. Una revisión de la estrategia para orientar la investigación hacia la resolución de problemas sociales, contenida en el Plan Institucional de Desarrollo 2015-2018, concluyó que la institución no cuenta con mecanismos para medir el impacto de los proyectos.
- VI. Es importante distinguir que, fundamentalmente, en la universidad se realizan evaluaciones con tres propósitos: a) promoción laboral y obtención de plazas, b) selección de propuestas para financiamiento y c) valoración de productos académicos comprometidos. En el primero de los casos, se utiliza un sistema cuantitativo de evaluación académica basado en el Reglamento de Ingreso y Promoción del Personal Académico de la Universidad el cual contempla distintas funciones para evaluar el desempeño de los profesores-investigadores. En torno de la actividad científica y de vinculación destacan por su valor los artículos científicos publicados en revistas indizadas o arbitradas, ser responsable o colaborador de proyectos con financiamiento (los cuales implican el compromiso de entregar productos académicos) y el registro de propiedad intelectual.
- VII. En el segundo de los casos, la Dirección de Investigación y Posgrado (DIP) y la Secretaría de Extensión Universitaria recurren a la dictaminación por pares para evaluar *ex ante* el valor de proyectos que concursan por financiamiento interno proveniente de distintos fondos. Finalmente, la evaluación *ex post* de los proyectos de vinculación e investigación registrados se concentra en torno de la entrega de productos académicos comprometidos. Estos pueden ser, entre otros, publicaciones, ponencias en congresos, registros de propiedad intelectual o la formación de recursos humanos (tesis).
- VIII. La Universidad ha perfeccionado durante años estos procedimientos de evaluación, y es innegable que aportan mucho valor a la organización, por ejemplo al transparentar los criterios en los que se basa la toma de decisiones. Sin embargo, es una buena práctica institucional evaluar los procedimientos con regularidad, especialmente cuando se han identificado elementos que podrían mejorar.
- IX. En concreto, el problema es que la productividad académica o el factor de impacto de las revistas no son útiles para rendir cuentas sobre los resultados obtenidos con las investigaciones aplicadas y los proyectos de vinculación. Si bien se espera que

los investigadores generen conocimientos y, por ello, parece lógico que su efectividad se mida con los artículos a través de los que difunden sus hallazgos, se ha señalado que ese indicador tiene múltiples problemas, especialmente cuando se utiliza como única medida de evaluación (Ekboir, 2016; Vasen y Lujano, 2017; Chavarro y Ràfols, 2018).

- X. Es un error pensar que las investigaciones son, en sí mismas, una prueba de efectos sociales positivos (Sivertsen y Meijer, 2020). Para tener un impacto, es necesario que el conocimiento generado se retome para incidir en las condiciones sociales, económicas o ambientales problematizadas (Netherlands Organization for Scientific Research, s.f.). En este sentido, los productos académicos no informan si el conocimiento se utilizó o no, y por lo tanto no hay forma de establecer su impacto.
- XI. En el fondo, si solo se utilizan indicadores ligados a la productividad académica, la evaluación se fundamenta en cifras y no en juicios que analizan el efecto del conocimiento generado (Hicks *et al.*, 2015). Como consecuencia, la evaluación se adhiere al principio “más es mejor” (Bensing *et al.*, 2013) y genera un sistema de incentivos que confunde los objetivos (resolver problemas sociales) con los indicadores utilizados (Campbell, 1976). Como resultado, incrementar el número de productos académicos puede convertirse en una meta en sí misma, pasando por encima de la relevancia del conocimiento.
- XII. Al resaltar la importancia de medir el impacto de las investigaciones, no se pretende cuestionar, en modo alguno, la relevancia de las investigaciones básicas. Éstas son fundamentales para generar contribuciones de largo plazo en nuestra base de conocimientos. Tampoco se pretende generar la impresión de que las investigaciones aplicadas deben tener impactos positivos siempre. La actividad científica es incierta por definición, se interactúa con múltiples actores y se prueban nuevas ideas todo el tiempo, por lo que fallar es un derecho que se da por descontado. En este contexto, las evaluaciones deben ser una herramienta de aprendizaje que permita a los investigadores ser cada vez más efectivos.

VISIÓN 2023

- XIII. Los criterios de evaluación basados en la productividad académica han sido fuertemente cuestionados a través de iniciativas como la Declaración de San Francisco sobre la Evaluación de la Investigación⁸⁴, el Manifiesto de Leiden sobre indicadores de investigación⁸⁵ y el Foro Latinoamericano sobre Investigación Científica (FOLEC-CLACSO)⁸⁶.
- XIV. En varios países de la Unión Europea, la comprensión de la brecha entre la producción de conocimiento y su impacto ha provocado un replanteamiento de la forma en la que se evalúa la práctica científica (Van der Meulen *et al.*, 2010; Spaapen *et al.*, 2011; Rafols *et al.*, 2017). Estas propuestas han buscado mejores maneras de conectar a los científicos con la sociedad y han sido sintetizadas en conceptos como transdisciplina, interacciones productivas, o intermediación de

⁸⁴ <https://sfdora.org/>

⁸⁵ <https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>

⁸⁶ <https://www.clacso.org/folec/>

conocimientos (*knowledge brokering*) (Spaapen y Van Drooge, 2011; Vallaey, 2014; Turnhout, 2018). En síntesis, estos enfoques valoran la ciencia por su orientación hacia la resolución de problemas concretos junto con otros actores que participan tanto en la definición del problema como en la creación de soluciones (Turnhout, 2018).

XV. Entre las prácticas recomendadas, se encuentran:

- Complementar los criterios cuantitativos utilizados para tomar decisiones de obtención de plazas y promoción laboral, de manera que, al valorar la trayectoria de un investigador, el tipo de revista o la cantidad de publicaciones sean sólo una variable dentro de una amplia gama de indicadores que reflejen la relevancia de su trabajo (Cagan, 2013). Sin embargo, en las valoraciones cualitativas es importante dejar claro y demostrar de forma pública cómo se atienden posibles sesgos, discriminación y otras formas de trato injusto (Science Europe, 2020).
- Promover enfoques híbridos de evaluación para que los científicos rindan cuentas y, simultáneamente, aprendan cómo ser más efectivos. Estas evaluaciones hacen énfasis en las interacciones entre equipos interdisciplinarios de investigadores y otros actores (Spaapen y Van Drooge, 2011), analizando la efectividad de los mecanismos de intercambio utilizados (LSE Public Policy Group, 2011). En específico, los indicadores seleccionados podrían reflejar: a) los medios a través de los cuales el conocimiento se dispersó entre los grupos interesados⁸⁷, b) el grado en que los grupos involucrados muestran interés y valoran los resultados de la investigación⁸⁸ o c) evidencia del uso e impacto del conocimiento⁸⁹ (Spaapen *et al.*, 2011).
- Evaluar las investigaciones en su propio contexto, utilizando indicadores *ad hoc* para cada disciplina científica (Hicks *et al.*, 2015). Esto es necesario pues no existe un modelo de evaluación que pueda aplicarse a todas las áreas de la ciencia.

XVI. Se desea que, para el año 2023, la Universidad sea un referente internacional en la promoción de la ciencia con relevancia social, pues gracias a su autonomía institucional ha redefinido la noción de impacto que está detrás de sus procedimientos de evaluación científica. Además, a nivel nacional la Universidad asumirá un rol de liderazgo para fomentar cambios en las políticas federales de evaluación científica, de forma que incentiven la obtención de impactos a través de investigaciones aplicadas.

XVII. Para conseguirlo, se requiere transformar la cultura de la investigación en la Universidad. Esta propuesta presenta una serie de actividades para promover cambios involucrando a la comunidad académica de la Universidad.

⁸⁷ Por ejemplo, a través de talleres multi-actor, páginas web o artículos de periódico.

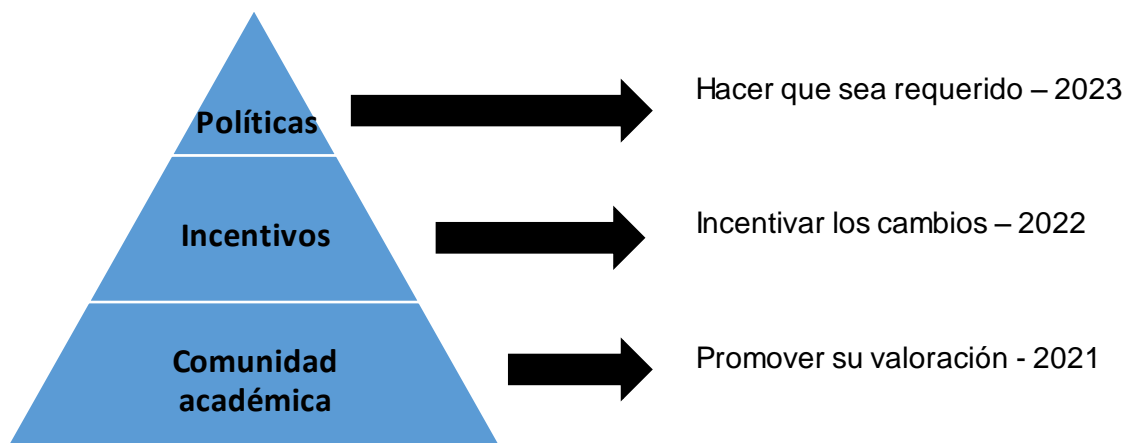
⁸⁸ Expresado, por ejemplo, a través de la co-producción de conocimientos, financiamiento o la invitación para formar parte de un consejo.

⁸⁹ Por ejemplo, con la adopción de prácticas, o la integración en reglamentos o políticas.

ACTIVIDADES

XVIII. Para organizar las actividades, la propuesta utiliza los niveles de intervención para cambiar una cultura de investigación (Figura 12) propuestos por Nosek (2019).

Figura 12. Niveles de intervención para cambiar una cultura de investigación.



Fuente: Modificado de Nosek (2019)

2021. Primera fase. Comunidad académica

- XIX. La Universidad firmará como organización la Declaración de San Francisco sobre la Evaluación de la Investigación: <https://sfdora.org/sign/>
- XX. Por facultad, se realizarán una serie de eventos utilizando técnicas del Arte del Liderazgo Participativo. Su objetivo será realizar encuentros multi-actor para:
- Reflexionar sobre la cultura de investigación de la Universidad, identificando el impacto que tiene en el comportamiento de los investigadores, en la calidad de los proyectos de investigación y en la sociedad.
 - Identificar desafíos, áreas de oportunidad y mejores prácticas en la evaluación de proyectos propios de las disciplinas pertenecientes a la Facultad.
 - Desarrollar estándares de calidad e indicadores *ad hoc* para la evaluación del impacto de los proyectos de investigación y vinculación de la Facultad.
- XXI. Se ofertará un curso de formación docente que brindará herramientas a los investigadores de la Universidad para el diseño de proyectos orientados a la obtención de resultados. En ese sentido, el curso incluirá la enseñanza de estrategias de intermediación de conocimientos y metodologías para evaluar su efectividad.

2022. Segunda fase. Sistema de incentivos

- XXII. La Dirección de Investigación y Posgrado y la Secretaría de Extensión Universitaria desarrollarán esquemas de financiamiento focalizado para atender desafíos tecnológicos, sociales y económicos específicos del estado. El esquema incluirá una interfaz que facilitará a los investigadores la integración de las teorías del cambio e indicadores de resultado de cada proyecto.
- XXIII. Los fondos anteriores funcionarán como un piloto para incorporar instrumentos de evaluación *ex ante* (dictaminación de propuestas) y *ex post* (evaluación de interacciones productivas). Al finalizar la implementación de los esquemas de financiamiento focalizado se hará un análisis de la robustez⁹⁰ de los procedimientos de evaluación utilizados.
- XXIV. Se desarrollará un sistema abierto de información que permitirá a la Universidad comunicar la pluralidad de impactos que genera a nivel estatal y nacional.

2023. Tercera fase. Ajustes en políticas

- XXV. Se revisará el sistema de evaluación académica basado en el Reglamento de Ingreso y Promoción del Personal Académico de la Universidad, con la finalidad de valorar la incorporación de indicadores que reconozcan el impacto generado por los investigadores.

⁹⁰ Es decir, valorar la capacidad de los instrumentos para evaluar de forma confiable la calidad e impacto de las propuestas.