

Las trampas de los «indicadores resbaladizos»: la importancia de leer entre líneas¹

JONATHAN FOX*

JULIA FISCHER-MACKEY**

Traducción: Karina Romera Scott

Recibido: 10 de agosto de 2022. Aceptado: 12 de octubre de 2022.

Resumen. En el ámbito de la control social, los estudios sobre «monitoreo comunitario» han elaborado amplias afirmaciones sobre «lo que funciona» —o no— en la práctica, con importantes implicaciones para los profesionales y los responsables políticos. La interpretación de estos resultados se complica cuando los estudios se basan en «indicadores resbaladizos» que no miden los procesos del mundo real que dicen abordar. Este artículo ilustra el problema de los indicadores resbaladizos, que tiene dos elementos principales. Primero, algunos estudios se basan en indicadores que no miden realmente el monitoreo o seguimiento comunitario. Segundo, los estudios que afirman que el seguimiento comunitario no ha conseguido mejoras, en realidad lo que pueden revelar es que no se ha llevado a cabo el seguimiento comunitario. Aunque la complejidad de los métodos de investigación puede ocultar estos dos problemas relacionados, los lectores pueden evaluar si las afirmaciones de los estudios están respaldadas por sus datos

.....
¹ Con autorización de la publicación original: Fischer-Mackey, J. y Fox, J. (2022): *Pitfalls of «slippery indicators»: the importance of reading between the lines*, *Development in Practice*, DOI: 10.1080/09614524.2022.2104220.

* Profesor en la Escuela de Servicio Internacional (SIS) de la American University, en Washington, D.C. Estados Unidos (correo electrónico: fox@american.edu | <https://orcid.org/0000-0002-2889-0397>).

** Profesora en la Escuela de Servicio Internacional (SIS) de la American University, en Washington, D.C. Estados Unidos (correo electrónico: juliafm@american.edu | <https://orcid.org/0000-0002-0336-3283>).

empíricos comprobando si los resultados se basan en indicadores que realmente miden lo que dicen estudiar.

Palabras clave: rendición de cuentas, participación, sociedad civil, gobernanza y políticas públicas.

Abstract. Within the field of social accountability, studies about «community monitoring» have made broad claims about «what works» —or not— in practice, with significant implications for practitioners and policymakers. Interpretation of these findings is complicated when studies rely on «slippery indicators» that do not measure the real-world processes they claim to address. This article illustrates the problem of slippery indicators, which has two main elements. First, some studies rely on indicators that do not actually measure community monitoring. Second, studies that claim to show a failure of community monitoring to deliver improvements may actually show a failure to deliver community monitoring in the first place. While complex research methods may obscure these two related problems, readers can still assess whether studies' claims are supported by their empirical data by checking whether the findings are grounded in indicators that actually measure what they claim to study.

Keywords: Accountability; participation; civil society; governance and public policy.

Introducción

En los últimos años se ha producido un flujo constante de estudios a gran escala sobre el monitoreo o seguimiento comunitario y otras intervenciones de control social que arrojan resultados nulos, publicados con titulares que sugieren que estos enfoques de reforma de la gobernanza «no funcionan».¹ Para los profesionales y los responsables de las políticas públicas, estos resultados pueden ser desconcertantes y, de hecho, desmoralizantes. Es cierto que muchos de estos esfuerzos fracasan, pero el problema de algunos de estos estudios es que no miden realmente los procesos del mundo real que pretenden abordar. Cuando los investigadores no definen los conceptos clave ni eligen indicadores que se

.....
¹ Estos estudios se centran en intervenciones que aprovechan el poder de la información para estimular e informar la acción cívica para mejorar el rendimiento del sector público. Véase, entre otros, Banerjee *et al.*, (2010); Björkman y Svensson (2009); Raffler, Posner y Parkerson (2019a, 2019b, 2020); Kosec y Wantchekon (2020); Lieberman, Posner y Tsai (2014).

ajusten a ellos, pueden acabar con «indicadores resbaladizos» que no miden lo que dicen medir. El problema de los «indicadores resbaladizos», que este artículo analiza en el contexto de los estudios de monitoreo comunitario, tiene dos elementos principales: primero, algunos estudios sacan conclusiones sobre el monitoreo comunitario, pero se basan en indicadores que no miden el monitoreo comunitario.

Segundo, los estudios que afirman que el seguimiento comunitario no ha conseguido mejoras en los servicios públicos, en realidad pueden revelar, en primer lugar, que no se ha llevado a cabo el monitoreo comunitario. Aun cuando la cuestión de la validez de la medición puede parecer un asunto técnico para los metodólogos, de hecho importa mucho a los profesionales y a los responsables políticos que buscan saber qué se puede hacer para avanzar en la rendición de cuentas. Este artículo muestra cómo los lectores no especializados pueden evaluar la «letra pequeña» de los estudios técnicos para apreciar si se basan en indicadores empíricos convincentes. Este enfoque permite a los lectores evaluar por sí mismos la credibilidad de las conclusiones de los encabezados.

El problema de los indicadores resbaladizos

¿Qué señalan realmente los indicadores? En la investigación social y la evaluación de programas, el término «indicador» se utiliza para referirse a una variable que sirve para medir el cambio relacionado con el objetivo de un programa o política. Los indicadores pueden observarse y medirse, y proporcionan información sobre conceptos más amplios que a menudo no pueden observarse o medirse directamente. Un concepto puede referirse a un proceso (como el monitoreo comunitario) o a un fenómeno (como la transparencia presupuestaria). Dado que la medición de los procesos de cambio en las relaciones entre los ciudadanos y el Estado es difícil, a menudo se necesitan múltiples indicadores empíricos para proporcionar una visión significativa de las cuestiones de gobernanza.

Un término técnico para evaluar «si una variable mide lo que se supone que mide» es la *validez de la medición*.² Una *medida válida* depende de dos cosas. Primero, requiere una definición explícita del concepto de interés (Adcock y Collier 2001; Sartori

.....

² Ver Bollen (1989: 184) and King, Keohane, and Verba (1994: 25), citado en Adcock and Collier (2001: 530).

1984)³ Esto es crucial porque los conceptos clave de interés para los científicos sociales pueden definirse de muchas maneras diferentes. Muchos conceptos en el campo de la gobernanza son ambiguos y controvertidos, como la participación, el empoderamiento, el monitoreo comunitario y la rendición de cuentas (por ejemplo, Alsop, Bertelson y Holland 2006; Fox 2022; Narayan 2005; Shutt y McGee 2013; VeneKlasen y Miller 2007). En el contexto de las múltiples definiciones y enfoques posibles para traducir estos conceptos en indicadores empíricos cuantificables, los investigadores deben explicitar las razones que justifican su elección de indicadores empíricos.

El segundo requisito para que una medida sea válida es que los indicadores empíricos capturen los elementos esenciales del concepto y no incluyan mucha información no relacionada a él. En pocas palabras, los indicadores deben decirnos algo significativo sobre la existencia de un concepto clave o un cambio en el mismo. Para explicar los motivos de su elección de indicadores, los investigadores deben abordar al menos dos cuestiones relacionadas.

Primero, ¿cómo captura un indicador un concepto? Segundo, ¿cuáles son los supuestos, los puntos fuertes y las limitaciones de ese indicador? Si la relación del indicador con el concepto es clara, lógica o está justificada por los investigadores, el lector puede confiar en él. Mientras que algunos investigadores pueden sentirse cómodos midiendo lo que suponen que es un indicador indirecto del proceso de interés, otros pueden no estar convencidos de esos supuestos. Si la relación del indicador con el concepto no es clara o es indirecta, los investigadores deben explicar a los lectores por qué es apropiado utilizar el indicador.

Si los investigadores no definen sus conceptos clave o no explican sus razones para seleccionar indicadores específicos, pueden acabar con lo que aquí llamamos «indicadores resbaladizos». Al igual que una sustancia resbaladiza difícil de manejar, los indicadores resbaladizos o escurridizos dificultan que los lectores sepan lo que realmente dicen sobre el concepto clave. El problema de los indicadores resbaladizos puede observarse en muchos tipos de investigación. Para ilustrar el problema, se analizan ejemplos procedentes principalmente de la literatura experimental. Aunque los experimentos de campo pueden

.....

³ También es importante que los indicadores no capten procesos que no estén relacionados con el concepto de interés. Según Sartori, los conceptos claros permiten a los analistas «distinguir A de lo que no es A» (1984: 74).

utilizar técnicas analíticas y de muestreo rigurosas, la validez de sus conclusiones depende de que sus pruebas empíricas subyacentes se basen en medidas convincentes de conceptos claramente definidos. En otras palabras, análisis técnicamente rigurosos de datos no implican, en primer lugar, que el proceso de interés se midió de forma convincente.⁴ Aquí es donde los lectores no técnicos pueden sacar sus propias conclusiones sobre la solidez del estudio, aunque no tengan la formación técnica necesaria para evaluar la aleatorización, las técnicas de muestreo o el análisis cuantitativo.

Los indicadores resbaladizos y los conceptos no especificados también generan confusión sobre cómo interpretar los resultados de la investigación empírica. Los profesionales quieren saber cuál es la conclusión. Sin embargo, dar sentido a las complejas conclusiones empíricas suele ser más una cuestión de interpretación y explicación de la lógica utilizada que una cuestión metodológica técnica. Este es un problema importante de la investigación en el amplio campo de la reforma de la gobernanza, porque algunos estudios afirman que las intervenciones no consiguen producir las mejoras esperadas en la prestación de servicios, cuando la evidencia presentada demuestra que las intervenciones en realidad no se llevaron a cabo.⁵ Esta cuestión de interpretación surge en los estudios de monitoreo

.....
⁴ Esta cuestión de si los indicadores captan el fenómeno de interés es distinta de la cuestión de si una intervención produjo resultados nulos porque era demasiado débil para hacer una diferencia —el problema de la «dosis baja». Véase Fox (2015) y Wein (2020).

⁵ Como ejemplo de la distancia entre los resultados de los titulares y las pruebas subyacentes, un importante estudio reciente de varios países afirmaba que la «policía de proximidad» (*community policing* en inglés) no había dado resultados, cuando las pruebas presentadas mostraban más bien un fracaso en la implantación de la policía comunitaria (Blair, 2021). Los autores de un experimento de campo que incluía la formación de los ciudadanos para profundizar la democracia local declararon el fracaso cuando sus pruebas etnográficas mostraban que en la práctica la intervención se realizaba de forma incompleta (Rao, Ananthpur y Malik 2017). Una prueba multinacional del poder de la información para fomentar la acción electoral también declaró resultados nulos, aunque las intervenciones fueron débiles y se cumplieron de forma incompleta (Dunning *et al.*, 2019). En otro ejemplo, un estudio que afirmaba que una intervención de transparencia y rendición de cuentas no había dado resultados, en realidad medía los efectos de una intervención de participación que implicaba poca transparencia o rendición de cuentas (Arkedis *et al.*, 2021).

comunitario, cuando los intentos de inducir el monitoreo comunitario se tratan como evidencia de que el monitoreo comunitario realmente ocurrió. Dado que los intentos de fomentar el monitoreo comunitario pueden funcionar o no, los investigadores deben proporcionar evidencia de que el monitoreo comunitario realmente ocurrió si van a afirmar que el monitoreo comunitario no «funcionó» para mejorar la prestación del servicio.⁶

El concepto de monitoreo comunitario dentro del campo del control social

El control social (*social accountability*) es una categoría general que incluye numerosos «mecanismos formales o informales a través de los cuales los ciudadanos y/o las organizaciones de la sociedad civil se empeñan en pedir cuentas a los funcionarios del Estado o a los proveedores de servicios» (Camargo y Jacobs, 2013: 7). El interés de los profesionales y los académicos por el control social creció rápidamente después de que el *Informe sobre el Desarrollo Mundial* de 2004 reconociera que la voz de los ciudadanos podía contribuir directamente a mejorar la prestación de servicios (Banco Mundial, 2004). Este «camino corto» hacia la rendición de cuentas sería, en teoría, más eficiente para mejorar los servicios que el largo proceso de la responsabilidad electoral (*electoral accountability*) y los subsiguientes esfuerzos de gestión.⁷

.....

⁶ Los autores de un experimento de campo que evalúa el efecto de la participación en la rendición de cuentas afirman explícitamente que el hecho de que las acciones de los ciudadanos para la rendición de cuentas tengan lugar realmente no es lo importante, sugiriendo que la amenaza potencial de dicha acción puede ser igualmente efectiva. Desde este punto de vista, se supone que proporcionar información relevante es tan poderoso como si los ciudadanos actuaran realmente: «...dado que tanto una protesta real como la amenaza de voz pueden haber desalentado a la élite política local de desviar recursos destinados a las escuelas, en equilibrio, no hay razón para creer que la incidencia de la voz y el desvío local de fondos deban estar correlacionados» (Reinikka y Svensson 2011: 959). La proposición es tratada como algo evidente y no como una hipótesis que necesita una justificación explícita y una prueba.

⁷ El análisis posterior distinguió entre respuestas y capacidad de respuesta a la voz de los ciudadanos. En la práctica, las respuestas oficiales a la voz de los ciudadanos pueden quedarse cortas, como cuando las mejoras en la prestación de servicios resultan ser puntuales, parciales o selectivas (Fox 2022).

Los esfuerzos de control social suelen implicar algún tipo de suministro o acceso a la información y el monitoreo de los servicios o acciones públicas, para informar la acción de los ciudadanos. Algunos enfoques del seguimiento se basan en el aumento de la transparencia o la difusión de datos oficiales, y otros implican la recopilación y el análisis de datos por parte de ONG técnicamente sofisticadas. Otro acercamiento a la producción y difusión de información consiste en el «monitoreo comunitario», que es la recopilación de información sobre los servicios públicos *por parte de los miembros de la comunidad* a la que están destinados esos servicios, con el fin de documentar los problemas y posiblemente hacer un seguimiento de las mejoras. Las actividades de monitoreo comunitario pueden adoptar muchas formas, pero siempre implican procesos inclusivos de generación y análisis de datos. Como explica Joshi «A menudo, el monitoreo comunitario se utiliza como una forma de garantizar que el desempeño en curso mantiene los estándares normales y se centra en características observables, por ejemplo, la asistencia de maestros o médicos, la calidad de la construcción en las instalaciones o que se sigan los procedimientos adecuados» (2013: 37).

Esta definición distingue entre recopilar datos y darles sentido (el «monitoreo comunitario»), y las posibles «acciones comunitarias» de seguimiento orientadas por los datos. Estas acciones pueden incluir llamar la atención del público o de los medios de comunicación sobre un problema, solicitar información a las autoridades, colaborar en la resolución de problemas con los funcionarios públicos, hacer que las autoridades rindan cuentas a través de recursos administrativos, plantear demandas a los funcionarios electos, o utilizar el sistema legal o la sanción social (por ejemplo, Hernández *et al.*, 2019; Fischer-Mackey *et al.*, 2020; Guertzovich y Poli 2020). Es importante señalar que la acción de la comunidad para mejorar la prestación de servicios también puede producirse sin monitoreo comunitario; del mismo modo, la recopilación y el análisis de datos pueden producirse sin una acción colectiva posterior. Por lo tanto, confundir el concepto de monitoreo comunitario con la voz y la acción posteriores complicaría los esfuerzos para identificar su contribución potencial. La información que es generada por los propios miembros de la comunidad puede ser más creíble para ellos y más relevante para sus necesidades (Flores, 2018). Los datos del monitoreo comunitario también pueden legitimar las preocupaciones de los miembros de la comunidad ante los proveedores de servicios y los funcionarios públicos (incluso si documentan problemas que ya eran

conocidos localmente, como el desabastecimiento de medicamentos o los patrones de falta de respeto y abuso).⁸

La conceptualización de otros autores sobre el monitoreo comunitario es mucho más amplia. Por ejemplo, una revisión sistemática incluye las campañas de divulgación externa que las ONG u otros utilizan para alentar a los miembros de la comunidad a tomar medidas como un tipo de monitoreo comunitario (Molina *et al.*, 2017: 5-6). Sin embargo, los esfuerzos para proporcionar información o alentar a las comunidades son posibles precursores que pueden o no conducir a la supervisión real por parte de las comunidades. La misma revisión también incluye los mecanismos de reparación de quejas (MRQ) como un *tipo* de monitoreo comunitario. Los MRQ son procesos institucionalizados que permiten a los ciudadanos presentar quejas a los gobiernos, para buscar reparación de los servicios que deberían haber recibido por derecho (Pande y Hossain 2022). Los MRQ no implican la recopilación de datos de la comunidad y sus conclusiones rara vez se hacen públicas. Proporcionan información, pero al gobierno, no a las comunidades. El punto aquí es que el monitoreo comunitario se refiere a un enfoque distinto al de la producción de evidencia, dentro de una amplia constelación de estrategias y tácticas de control social. Definir todas las herramientas para la producción o provisión de información local como monitoreo comunitario, o tratar el concepto como intercambiable con la amplia categoría general de control social, complica los esfuerzos para identificar las fortalezas y limitaciones que son específicas del monitoreo comunitario.

En resumen, el monitoreo comunitario puede adoptar muchas formas, pero en su esencia, el concepto consiste en la recopilación de información sobre los servicios públicos por parte de los *miembros de una comunidad afectada* con el fin de documentar los problemas y apoyar las acciones para abordarlos. A continuación, se presentan breves ejemplos de estudios que afirman la existencia de un monitoreo comunitario sin medir realmente dichas acciones.

.....

⁸ El tipo de resultados del monitoreo que generan credibilidad ante las autoridades locales puede ser diferente del tipo de evidencia que pretende cumplir con los estándares de las ciencias sociales. Flores (2018) describe cómo los monitores comunitarios de base recopilaron el testimonio de los usuarios en lugar de los datos cuantitativos convencionales sobre los problemas del sistema de salud porque era más influyente con las autoridades locales.

Tabla 1 | Indicadores resbaladizos de «monitoreo comunitario»

- (1) Buzones de sugerencias
- (2) Fichas de espera numeradas
- (3) Cartel informativo de servicios gratuitos
- (4) Cartel informativo de los derechos del paciente
- (5) *Discutir la clínica en reuniones [del consejo local]*
- (6) Recibir información acerca de [Comité de Administración de Unidad de Salubridad]

Fuente: Björkman and Svensson (2009, 749, tabla 2)

Indicadores de seguimiento comunitario resbaladizos

«Power to the People» (Björkman y Svensson, 2009) fue un influyente experimento de campo que proporcionó a las comunidades de Uganda fichas informativas con datos sobre la utilización de los centros de salud y la percepción de la calidad de los servicios. El estudio pretendía evaluar si la información sobre los servicios sanitarios incitaría a la gente a actuar de forma que mejoraran los servicios sanitarios y los resultados de salud. El estudio encontró una mejora de los resultados sanitarios y la atribuyó al «monitoreo comunitario». Aunque el estudio no definió lo que entendía por «monitoreo comunitario», operacionalizó el concepto con un índice que incluía seis indicadores (véase la tabla 1). De los seis indicadores, sólo uno estaba relacionado con el concepto de seguimiento comunitario, tal y como ha sido discutido anteriormente.

Este índice no refleja ninguna definición coherente de monitoreo comunitario y confunde procesos vagamente relacionados y no relacionados con el monitoreo comunitario. Los autores no explican por qué estos indicadores deben tomarse como prueba de los procesos o resultados del monitoreo comunitario. Tres de los seis indicadores (3, 4, 6) representan la diseminación de información unidireccional a las comunidades, pero no sobre el desempeño de la clínica. Los indicadores 1 y 2 representan herramientas que proporcionan información a los administradores de las clínicas. El estudio no especifica por qué o cómo pueden estar relacionados con la recopilación o el uso de datos por parte de la comunidad. Sólo el indicador 5 (si los miembros de la comunidad hablaron del centro de salud en las reuniones comunitarias) podría indicar un seguimiento por parte de la comunidad, pero incluso en ese caso, no se explica la justificación de tal suposición. A pesar de esta dilución del monitoreo comunitario real en el índice, que es la base del aná-

lisis cuantitativo, el estudio también incluye evidencia adicional —no incluida en el índice cuantitativo— que ofrece una explicación mucho más plausible de la mejora en la prestación de servicios. Antes de la intervención, los comités oficiales de gestión de los centros de salud se consideraban en general ineficaces. En las zonas de tratamiento, un tercio de esos comités se reconstituyeron o eligieron nuevos dirigentes, en contraste con la ausencia de tales cambios en el grupo de control (Björkman y Svensson 2009: 747). Estos cambios de liderazgo en los organismos oficiales encargados de la supervisión son indicadores convincentes de que se ha reforzado la vigilancia comunitaria. Por el contrario, los datos sobre la presencia de carteles en las clínicas son indicadores de campañas informativas (como se ha señalado anteriormente) más que de monitoreo comunitario.

Un estudio de seguimiento relacionado con «Power to the People», realizado por Raffer y sus colegas (2019 a, 2019 b, 2020), intentó replicar los hallazgos generales del estudio original a mayor escala. Sin embargo, no encontraron impactos positivos en los resultados de salud.⁹ Raffer y colegas afirmaron que documentaron lo que llamaron monitoreo comunitario, pero incluyeron pocos indicadores de monitoreo comunitario. Sólo uno de los cuatro componentes del índice de Raffer, Posner y Parkerson (2019 a) puede haber sido un indicador razonable de la supervisión de la comunidad: una pregunta sobre si la clínica se discutió en las reuniones de la comunidad (véase la tabla 2).

Como se muestra en la tabla 2, una de las cuatro preguntas del índice de monitoreo comunitario del estudio (1) se limitaba a preguntar por la asistencia a las reuniones comunitarias (un indicador débil de la participación, pero no del monitoreo). Otras dos (3 y 4) preguntas sólo se referían a escenarios hipotéticos, en lugar de preguntar si el monitoreo comunitario se produjo realmente y, en ese caso, si los proveedores de servicios respondieron y cómo lo hicieron (2019 a, anexo A, 40). Una medida mucho más directa del monitoreo comunitario habría abordado si realmente documentaron el absentismo.

.....

⁹ El borrador público del estudio no informa los componentes específicos del boletín de calificaciones, más allá de las referencias generales a la utilización y las percepciones de calidad y satisfacción. No se discute la relevancia percibida de estos indicadores para los ciudadanos. Al parecer, los boletines de notas no incluían indicadores más específicos y tangibles del rendimiento de las clínicas, como las tasas de desabasto de medicamentos, el absentismo del personal o los puestos no cubiertos.

Tabla 2 | Indicadores resbaladizos de «monitoreo comunitario»

(7) «Si los miembros del hogar informan que han asistido a las reuniones [del consejo local] en el último año».

(8) «*Si los miembros del hogar que asistieron a la reunión [del consejo local] informan de que se habló acerca del centro de salud local.*

(9) «Si los miembros de la familia *piensan* que los miembros de la comunidad comprometidos se enterarían si un trabajador sanitario no hiciera el esfuerzo que debería hacer para atender a sus pacientes» (énfasis añadido).

(10) «Si los miembros de la familia *piensan* que los miembros de la comunidad comprometidos se enterarían si un trabajador sanitario no se presentara a trabajar» (énfasis añadido).

Fuente: Raffler, Posner y Parkerson (2020, A3).

El estudio no justifica de su dependencia de indicadores muy indirectos que confunden las preguntas hipotéticas sobre percepción de la agencia con las pruebas de monitoreo real de la comunidad.

Los autores afirman que «tampoco encontramos pruebas de que los ciudadanos hayan aumentado su monitoreo o sanción a los trabajadores de la salud» (Raffler, Posner y Parkerson 2020, página del título). Sin embargo, los datos presentados no incluyen una medida directa de la monitoreo de la comunidad, por lo que no está claro cómo podrían saber los autores si éste se produjo. Además, a pesar de su interpretación de que su intervención no produjo presión ciudadana, continúan afirmando que «En contra de la literatura, el vínculo entre la provisión de información y el comportamiento de los proveedores [de salud] no pasó por la presión ciudadana» (Raffler, Posner y Parkerson 2020: 25); en otras palabras, afirman que la presión ciudadana no conduce a la mejora de los servicios. No explican la lógica de su conclusión de que la presión ciudadana no influye en la prestación de servicios, al tiempo que afirman no tener pruebas de que el monitoreo comunitario se haya producido. Cabe destacar que, una vez concluido el ensayo controlado aleatorizado y eliminadas las limitaciones de los protocolos del estudio, el mismo donantes del proyecto apoyó una campaña de derechos sanitarios que utilizó el monitoreo comunitario de los servicios de salud para informar la vocalización de demandas en múltiples niveles del sistema, al tiempo que supervisaba las tendencias de respuesta del gobierno (Bailey y Mujune, 2021). Tanto el estudio original de Björkman y Svensson (2009) como el de

seguimiento de Raffler, Posner y Parkerson (2019a: 2020) no definen el monitoreo comunitario, incluyen indicadores débiles del mismo y complican la interpretación de sus datos al incluir medidas de fenómenos no relacionados en sus índices.

Otro ejemplo notable de indicadores resbaladizos se encuentra en el sector educativo: «Pitfalls of Participatory Programs: Evidence From a Randomised Evaluation in Education in India» (Banerjee *et al.*, 2010). El estudio se propuso comprobar si «la participación de los beneficiarios en el monitoreo de los servicios públicos» (Banerjee *et al.*, 2010: 1) podía servir de base para la adopción de medidas destinadas a mejorar los resultados educativos. Sin embargo, el estudio no encontró pruebas de que se produjera una supervisión comunitaria de las escuelas. En cambio, la intervención incluyó el monitoreo de los resultados de aprendizaje de los alumnos y un intento infructuoso de activar los comités oficiales de supervisión de las escuelas.

En primer lugar, los voluntarios de la comunidad recopilaron datos sobre el aprendizaje de los alumnos mediante la realización de evaluaciones individuales a los niños, encontrando bajos niveles de aprendizaje. Sin embargo, los bajos resultados de aprendizaje pueden atribuirse a una serie de problemas sociales, económicos, personales e institucionales (por ejemplo, problemas emocionales, problemas de aprendizaje, baja asistencia a la escuela, hambre y estrés, desajuste entre la lengua de enseñanza y la lengua hablada en casa, ausentismo de los profesores, baja calidad de la enseñanza, etc.). Por eso, datos sobre el aprendizaje de los niños no son un sustituto de los datos sobre los servicios públicos en las escuelas como instalaciones y útiles escolares, libros de texto, calidad y asistencia de los profesores, *curriculum* y otros indicadores que se relacionan directamente a la prestación de servicios educativos. Además, el proceso de recopilación y análisis de datos sobre los servicios públicos por parte de los miembros de la comunidad puede ser un paso importante en el desarrollo de un plan para abordar los problemas. Por lo tanto, el carácter individual de los voluntarios que examinan a los niños es fundamentalmente diferente del concepto básico de monitoreo comunitario de prestación de servicios.

En segundo lugar, el estudio se basó en la idea de que la información sobre los malos resultados de aprendizaje de los alumnos llevaría a los miembros de la comunidad y a los organismos oficiales de supervisión de la educación (Comités Comunitarios de Educación, o CCE) a mejorar la calidad de la educación impartida en las escuelas de la villa. Los CCE son órganos locales designados, compuestos por el director de la escuela, el jefe del

gobierno de la aldea y tres padres de familia elegidos por los funcionarios del gobierno. En principio, los CCE pueden supervisar a los profesores, presentar quejas sobre determinados profesores a las autoridades educativas, contratar más profesores auxiliares, asignar recursos mínimos y solicitar al gobierno más recursos. El único control de los servicios públicos que podría haber tenido lugar según el diseño del programa habría sido el de los CCE. Sin embargo, el estudio no indicaba si los CCE o la comunidad disponían de datos relevantes sobre el rendimiento de las escuelas (como el absentismo de los profesores o la disponibilidad de libros de texto).

Antes de la intervención, un amplio estudio de referencia descubrió que los CCE sólo existían sobre el papel (Banerjee *et al.*, 2010: 13-14). Sin embargo, incluso después de las intervenciones destinadas a activar los comités de supervisión, sólo el siete por ciento de los padres de familia declaró saber que los CEV existían (Banerjee *et al.*, 2010: 22). Esto, junto con la ausencia de indicadores relacionados, sugiere que los CEV no participaban activamente en la supervisión de las escuelas.

El punto de partida del estudio era que los «beneficiarios» debían «tener la información necesaria para supervisar a los proveedores» (Banerjee *et al.*, 2010: 2). Sin embargo, su concepto del «monitoreo comunitario» se centró en cambio en el aprendizaje individual de los alumnos. Por lo tanto, no está claro si de hecho los miembros de la comunidad culpaban del mal aprendizaje de los alumnos al sistema escolar o a los propios alumnos, a la pobreza o a algún otro factor. El título del artículo «Los escollos de los programas participativos» sugiere que el estudio puso a prueba un programa que era participativo pero que no consiguió las mejoras esperadas en la prestación de servicios. En cambio, el estudio documentó otra cosa: una intervención que, en primer lugar, no logró inducir la supervisión de la comunidad, y por lo tanto no probó el impacto del monitoreo comunitario en la prestación de servicios.

Mejores indicadores de monitoreo comunitario

Un ejemplo de medidas sólidas de monitoreo comunitario puede encontrarse en un estudio de iniciativas lideradas por ciudadanos para el derecho a la salud en Guatemala (Hernández *et al.*, 2019). En el estudio se informa sobre dos tipos de monitoreo comunitario. El primero consistía en la recopilación de datos por parte de los ciudadanos sobre los servicios de los centros de salud. Para efectos de un análisis cualitativo comparativo,

el monitoreo de los establecimientos de salud fue calificado de acuerdo al: «[n]ivel de participación y continuidad en la recolección de evidencias a través de visitas a los establecimientos de salud, entrevistas a usuarios, asambleas comunitarias, [y] plataforma de reclamos por SMS», asignándose el mayor puntaje a los ejemplos en los que «las actividades de monitoreo regular [ocurrieron] por más de un año, generando gran cantidad de evidencias» (Hernández *et al.*, 2019: 395).

El segundo tipo de monitoreo se refería a la interacción y trabajo de supervisión ciudadana con las autoridades por parte de los defensores voluntarios del derecho a la salud y de las ONG asociadas, que incluía los progresos realizados hacia los resultados deseados por los ciudadanos. La interlocución con las autoridades se evaluó sobre la base de

La frecuencia y el enfoque de las reuniones con las autoridades municipales, sanitarias de distrito y de nivel superior, incluidas las autoridades provinciales y nacionales, y las instituciones de derechos humanos; señales de seguimiento y apoyo activo (por ejemplo, provisión de espacio para reuniones) por parte de las autoridades municipales, sanitarias de distrito y de nivel superior; interacción frecuente con las autoridades municipales centrada en la acción y el seguimiento de los problemas presentados; e interacción frecuente con las autoridades de salud del distrito con enfoque en la acción y seguimiento de los problemas presentados; y audiencia con dos o más autoridades de nivel superior o múltiples interacciones para dar seguimiento al problema presentado (Hernández *et al.*, 2019: 395).

Estos indicadores empíricos del monitoreo comunitario documentan el monitoreo comunitario que *realmente* ocurrió, no simplemente que se realizaron esfuerzos para promover el monitoreo. Además, documentan el seguimiento de las acciones ciudadanas que se basan en los datos del seguimiento, las que después pueden analizarse para informar las estrategias de control social. Este ejemplo demuestra cómo evitar el problema de los indicadores resbaladizos proporcionando a los lectores una comprensión precisa de lo que el estudio realmente observó y de las pruebas de sus conclusiones

Discusión y conclusiones

Este artículo ilumina el problema de los «indicadores resbaladizos» de los estudios que analizan el monitoreo y seguimiento de la comunidad para mejorar la prestación de servi-

cios públicos. Este nuevo concepto de indicadores resbaladizos nombra una desconexión entre lo que afirman los investigadores en sus conclusiones principales y lo que miden realmente sus medidas.

Este debate indica que incluso los notables experimentos de campo, que son rigurosos en su aleatorización, selección de muestras y análisis de datos, pueden no estar basados en medidas empíricas convincentes de los conceptos clave que estudian. Aun cuando los profesionales y los responsables políticos no estén en condiciones de evaluar el aspecto técnico del análisis de los datos, sí están en condiciones de evaluar si los investigadores hacen suposiciones convincentes sobre sus indicadores. Para que los lectores tomen decisiones informadas sobre cómo evaluar los resultados de la investigación, los investigadores deben ser más coherentes y explícitos sobre cómo definen los conceptos clave y sobre la justificación de cómo se miden empíricamente.

Los lectores interesados pueden suponer que los procesos de dictamen externo de las revistas académicas garantizan que las interpretaciones de los investigadores se ajustan a sus resultados empíricos reales. El problema de los indicadores resbaladizos —al menos en los ámbitos de la gobernanza y la reforma sectorial— sugiere que los lectores no pueden dar por sentada esa coherencia. Los patrones de desconexión ilustrados aquí, en los que los titulares y los resúmenes resultan no estar basados en las pruebas, sugieren que vale la pena que el lector revise las definiciones de los autores en relación a los conceptos y los indicadores empíricos, y se pregunte: ¿Cómo se justificaron esas elecciones de investigación? ¿Indican realmente los indicadores lo que afirman? En este artículo se han identificado varios casos en los que las conclusiones de los titulares se basan en suposiciones no examinadas sobre cómo interpretar los datos subyacentes.

La cuestión de si los indicadores indican realmente lo que los investigadores afirman se plantea de modo central cuando se trata de afirmaciones sobre «lo que funciona». En los estudios sobre gobernanza y reforma sectorial, las conclusiones de los autores, los resúmenes y los títulos de los artículos suelen desplegar mensajes de autoridad formulados en términos de si determinados tipos de intervenciones de transparencia, participación y rendición de cuentas «funcionan», o no. Sin embargo, esta discusión muestra que las evaluaciones de los investigadores sobre «lo que funciona» pueden ser una cuestión de interpretación, especialmente si las intervenciones encontraron obstáculos en su diseño o implementación.

Para los profesionales, *cómo* llevar a cabo la reforma de la gobernanza es una preocupación central y sin embargo, formulaciones en términos de «si funciona» antes de nada alejan a los investigadores del reto de cómo llevar a cabo la reforma de la gobernanza. Para ser convincentes, los estudios que pretenden evaluar si el control social «funciona» deben demostrar que esos procesos ocurrieron realmente en la práctica. Para resumir esta cuestión de cómo interpretar los resultados empíricos, tanto para los profesionales como para los analistas, hay una gran diferencia entre las reformas de la gobernanza que no funcionan y el fracaso de las reformas de la gobernanza. El primer lugar, donde los lectores informados deben buscar este tipo de deslizamiento entre los titulares y los resultados es en los indicadores empíricos.

En conclusión, la principal contribución de este artículo es introducir el concepto de indicadores resbaladizos y animar a los lectores no especializados a buscarlos para evaluar la credibilidad de las investigaciones que encuentran.

Referencias

- Adcock, R., y D. Collier (2001). «Measurement Validity: A Shared Standard for Qualitative and Quantitative Research». *American Political Science Review*, 95(3): 529-546.
- Alsop, R., M. Bertelson, y J. Holland (2006). *Empowerment in Practice: From Analysis to Implementation*. Washington, DC: World Bank.
- Arkedis, J., J. Creighton, A. Dixit, A. Fung, S. Kosack, D. Levy y C. Tolmie (2021). «Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania». *World Development*, 142: 105369.
- Bailey, A., y V. Mujune (2021). «Bottom-Up Accountability in Uganda: Learning from People-Centered, Multi-Level Health Advocacy Campaigns». Working Paper No. 8. Washington, DC: Accountability Research Center. <https://accountabilityresearch.org/publication/bottom-up-accountability-in-uganda-learning-from-people-centered-multi-level-health-advocacy-campaigns/>
- Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster y S. Khemani (2010). «Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India». *American Economic Journal: Economic Policy*, 2(1): 1-30.

- Björkman, M., y J. Svensson (2009). «Power to the People: Evidence From a Randomized Field Experiment on Community-Based Monitoring in Uganda». *Quarterly Journal of Economics*, 124(2): 735-769.
- Blair, G. (2021). «Community Policing Does Not Build Citizen Trust in Police or Reduce Crime in the Global South». *Science*, 374(6571): 1-14.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Camargo, C. B., y E. Jacobs (2013). «Social Accountability and its Conceptual Challenges: An Analytical Framework». Working Paper 16. Basel: Basel Institute on Governance. <https://edoc.unibas.ch/66317/>.
- Dunning, T., G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh y G. Nellis (2019). *Information, Accountability and Cumulative Learning: Lessons from Metaketa I*. Cambridge: Cambridge University Press.
- Fischer-Mackey, J., B. Batzin, P. Culum y J. Fox. (2020). «Rural Public Health Systems and Accountability Politics: Insights from Grassroots Health Rights Defenders in Guatemala». *Journal of Peasant Studies*, 47(5): 899-926.
- Flores, W. (2018). «How Can Evidence Bolster Citizen Action? Learning and Adapting for Accountable Public Health in Guatemala». Accountability Note 2. Washington, DC: Accountability Research Center. <https://accountabilityresearch.org/publication/how-can-evidence-bolster-citizen-action-learning-and-adapting-for-accountable-public-health-in-guatemala/>.
- Fox, J. (2015). «Social Accountability: What Does the Evidence Really say?» *World Development*, 72: 346-361.
- (2022). «Accountability Keywords». Working Paper 11. Washington, DC: Accountability Research Center. <https://accountabilityresearch.org/publication/accountability-keywords/>.
- Guerzovich, M. F. y M. Poli (2020). How Social Accountability Strengthens Cross-Sector Initiatives to Deliver Quality Health Services? Global Partnership for Social Accountability Note 17. Washington, DC: The World Bank. <https://documents1.worldbank.org/curated/en/600891606911830725/pdf/How-Social-Accountability-Strengthens-Cross-Sector-Initiatives-to-Deliver-Quality-Health-Services.pdf>.
- Hernández, A., A. L. Ruano, A. Hurtig, I. Goicolea, M. San Sebastián y W. Flores (2019). «Pathways to Accountability in Rural Guatemala: A Qualitative Comparative Analy-

- sis of Citizen-Led Initiatives for the Right to Health of Indigenous Populations». *World Development*, 113: 392-401.
- Joshi, A. (2013). «Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery». *Development Policy Review* 31: s29-s48. King, G., R. O. Keohane y S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Kosec, K. y L. Wantchekon (2020). «Can Information Improve Rural Governance and Service Delivery?» *World Development*, 125: 1-13.
- Lieberman, E., D. Posner y L. Tsai. (2014). «Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya». *World Development*, 60: 69-83.
- Molina, E., L. Carella, A. Pacheco, G. Cruces y L. Gasparini (2017). «Community Monitoring Interventions to Curb Corruption and Increase Access and Quality in Service Delivery: A Systematic Review». *Journal of Development Effectiveness*, 9(4): 462-499.
- Narayan, D. (2005). «Measuring Empowerment: Cross-Disciplinary Perspectives». Washington, DC: The World Bank.
- Pande, S. y N. Hossain (2022). «Grievance Redress Mechanisms in the Public Sector A Literature Review». Washington, DC: Open Governance Partnership and Accountability Research Center. <https://www.opengovpartnership.org/documents/grievance-redress-mechanisms-in-the-public-sector-a-literature-review/>.
- Raffler, P., D. Posner y D. Parkerson (2019a). «The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector». *Innovations for Poverty Action* (March 26 Draft) <https://www.povertyaction.org/sites/default/files/publications/The%20Weakness%20of%20Bottom-Up%20Accountability.pdf>.
- Raffler, P., D. Posner y D. Parkerson (2019b). «Does Bottom-Up Accountability Work? Evidence from Uganda». *VoxDev*. July 22. <https://voxdev.org/topic/health-education/does-bottom-accountability-work-evidence-uganda>.
- , ——— y ——— (2020). «Can Citizen Pressure be Induced to Improve Public Service Provision?» *Innovations for Poverty Action* (Oct. 30 draft). https://www.poverty-action.org/sites/default/files/publications/Citizen%20Pressure_Oct_302020.pdf.
- Rao, V., K. Ananthpur y K. Malik (2017). «The Anatomy of Failure: An Ethnography of

- a Randomized Trial to Deepen Democracy in Rural India». *World Development* 99: 481-497.
- Reinikka, R. y J. Svensson (2011). «The Power of Information in Public Services: Evidence from Education in Uganda». *Journal of Public Economics*, 95: 956-966.
- Sartori, G. (1984). *Social Science Concepts: A Systematic Analysis*. Thousand Oaks, CA: Sage.
- Shutt, C. y R. McGee. (2013). «Improving the Evaluability of INGO Empowerment and Accountability Programmes». Practice Document. No. 1. Sussex: Centro para el Impacto del Desarrollo. <https://www.ids.ac.uk/publications/improving-the-evaluability-of-ingo-empowerment-and-accountability-programmes/>.
- VeneKlasen, L. y V. Miller, eds. (2007). «A New Weave of Power, People and Politics: The Action Guide for Advocacy and Citizen Participation». Rugby: Practical Action Publishing.
- Wein, T. (2020). «Intervening Intensely: How to Estimate Dosage in Accountability Interventions». April 23. https://medium.com/@tom_wein/intervening-intensely-how-to-estimate-dosage-in-accountability-interventions-1da1b51d051a.
- World Bank (2004). *World Development Report 2004: «Making Services Work for Poor People»*. Washington, DC: World Bank.