

Criterios de evaluación

Greet Peersman

CENTRO DE INVESTIGACIONES INNOCENTI DE UNICEF

El Centro de Investigaciones Innocenti es la oficina de UNICEF especializada en investigación científica. El objetivo principal del Centro de Investigaciones es mejorar la comprensión internacional de una serie de cuestiones relacionadas con los derechos de la infancia, a fin de facilitar la plena aplicación de la Convención sobre los Derechos del Niño en todo el mundo. El centro tiene el cometido de establecer un marco integral de investigación y conocimiento dentro de la organización para brindar apoyo a los programas y políticas mundiales de UNICEF, y trabaja con los asociados para formular políticas con base empírica en favor de la infancia. Las publicaciones elaboradas por el centro contribuyen al debate global sobre la infancia y los derechos del niño e incluyen una amplia gama de opiniones.

Las opiniones expresadas corresponden a los autores o editores y se publican para estimular un mayor diálogo sobre métodos de análisis de impacto. Esta publicación no refleja necesariamente las políticas o perspectivas de UNICEF.

SINTESIS METODOLOGICAS DEL CENTRO DE INVESTIGACIONES

Las síntesis metodológicas del Centro de Investigaciones de UNICEF pretenden compartir prácticas de investigación, métodos, diseños y recomendaciones de reconocidos investigadores y analistas. La audiencia a la que van dirigidas es principalmente el personal de UNICEF que lleve a cabo, encargue o interprete los resultados de investigación y análisis para la toma de decisiones sobre programas, políticas y actividades de sensibilización.

Esta síntesis metodológica ha seguido un proceso de revisión por pares interna.

El texto no ha sido editado de acuerdo con los estándares de publicación oficiales y UNICEF declina toda responsabilidad por posibles errores.

Se permite la reproducción de cualquier parte de la presente publicación siempre que se incluya referencia a la presente. Si se desea utilizar una parte sustancial o la totalidad de la publicación dirijan su solicitud al Departamento de Comunicación en la dirección de correo electrónico: florence@unicef.org

Para consultas o descargas, pueden encontrar estas síntesis metodológicas en <http://www.unicef-irc.org/KM/IE/>

Recomendamos la siguiente cita para cualquier referencia al presente documento:

Peersman, G. (2014). Criterios de evaluación, *Síntesis metodológicas: evaluación de impacto* n.º 3, Centro de Investigaciones de UNICEF, Florencia.

Agradecimientos: Varios autores han proporcionado orientación en la preparación de esta síntesis. El autor y el Centro de Investigaciones de UNICEF desean agradecer a todos aquellos que han participado en la preparación de la presente publicación, especialmente a:

Por su contribución: Dugan Fraser, Patricia Rogers

Por su revisión: Nikola Balvin, Samuel Bickel, David Parker

© Fondo de las Naciones Unidas para la Infancia (UNICEF), septiembre de 2014

Centro de Investigaciones Innocenti de UNICEF

Piazza SS. Annunziata, 12

50122 Florencia (Italia)

Tel.: (+39) 055 20 330

Fax: (+39) 055 2033 220

florence@unicef.org

www.unicef-irc.org

1. CRITERIOS DE EVALUACIÓN: BREVE DESCRIPCIÓN

La evaluación se basa en una combinación de hechos y valores (principios, atributos o cualidades que se consideran intrínsecamente buenos, deseables, importantes y de utilidad general¹, por ejemplo «ser justos con todos») para calibrar el mérito de una intervención (es decir, de un programa o una política). Los criterios de evaluación especifican los valores que se emplearán en esta. Si bien pueden emplearse en distintos tipos de evaluaciones, esta síntesis aborda concretamente su uso en las evaluaciones de impacto (estudios que brindan información sobre los efectos a largo plazo que genera una intervención; véase la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)).

Las evaluaciones de UNICEF recurren a diversos criterios de evaluación. No todos se aplican en todas las evaluaciones, dado que algunos solo resultan adecuados para determinadas intervenciones o tipos de evaluación. Los [términos de referencia](#) de la evaluación deben señalar qué criterio es de aplicación.

Estos son algunos de los criterios de evaluación que rigen las evaluaciones de UNICEF sobre el impacto de las intervenciones²:

Los criterios habituales del Comité de Asistencia para el Desarrollo de la Organización de Cooperación y Desarrollo Económicos³

- **Pertinencia:** Medida en que los objetivos de la intervención son congruentes con los requisitos de los beneficiarios, las necesidades del país, las prioridades globales y las políticas de los asociados.
- **Eficacia:** Medida en que se lograron o se espera lograr los objetivos de la intervención, tomando en cuenta su importancia relativa.
- **Eficiencia:** Medida en que los recursos/insumos (fondos, especialización, tiempo, equipos, etc.) se han convertido económicamente en resultados.
- **Impacto:** Efectos de largo plazo positivos y negativos, primarios y secundarios, producidos directa o indirectamente por la intervención, intencionalmente o no.
- **Sostenibilidad:** Continuación de los beneficios de la intervención después de que haya cesado la principal asistencia para el desarrollo. Las intervenciones han de ser sostenibles desde una perspectiva ambiental y financiera. Cuando el énfasis no se sitúa en la ayuda externa, cabe definir la sostenibilidad como la capacidad de los principales interesados para sostener los beneficios de una intervención, una vez que se interrumpe la financiación de los donantes, con medidas que utilizan los recursos disponibles en la zona.

Criterios sobre la equidad, la igualdad de género y la adopción de un enfoque de programación basado en los derechos humanos

- **Equidad:** Justicia esencial de los procesos de decisión y sus resultados. Para UNICEF, implica que todos los niños tienen la oportunidad de sobrevivir, desarrollarse y alcanzar su pleno potencial, sin ser víctimas de la discriminación, los prejuicios o el favoritismo.

¹ Stufflebeam, Daniel L., «Evaluation values and criteria checklist», Western Michigan University Checklist Project, Kalamazoo, 2001. Véase https://www.wmich.edu/sites/default/files/attachments/u350/2014/values_criteria.pdf.

² Cuando existan divergencias importantes en la aplicación de los criterios de evaluación de un programa o una política, se mencionarán expresamente en esta síntesis.

³ Organización de Cooperación y Desarrollo Económicos, Comité de Asistencia para el Desarrollo, «Evaluation of development programmes, DAC Criteria for Evaluating Development Assistance», página web, OCDE. Véase <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

- **Igualdad de género:** Para UNICEF, implica promover la igualdad de derechos de las mujeres y las niñas, y apoyar su participación plena en el desarrollo político, social y económico de sus comunidades⁴.
- **Enfoque de programación basado en los derechos humanos:** Se sustenta en cinco principios rectores que pueden aplicarse como criterios de evaluación: la normatividad, la no discriminación, la participación, la transparencia y la rendición de cuentas⁵.

Criterios de evaluación⁶ para la asistencia humanitaria (se aplican en combinación con alguno de los anteriores, incluso con todos ellos)

- **Cobertura:** Necesidad «de llegar a grupos importantes de población que hacen frente a padecimientos que ponen en riesgo su vida, dondequiera que estén, a fin de brindarles una asistencia y una protección adecuadas a sus necesidades y desprovistas de agendas políticas ajenas»⁷.
- **Coordinación:** Necesidad de que los agentes que participan en la respuesta de emergencia se coordinen entre sí. La intervención de un único organismo no puede evaluarse de manera independiente, ya que una actuación que se considera adecuada desde el punto de vista de un agente puede no serlo desde la perspectiva del sistema en su totalidad.
- **Protección:** La provisión de seguridad y protección a la población beneficiaria (por ejemplo, la protección frente a elementos armados que operan en la zona de un proyecto, o los servicios de seguridad en un campamento de desplazados o refugiados) es fundamental para la eficacia de la acción humanitaria o las operaciones de socorro.
- **Coherencia** (es decir, coherencia de las políticas): Necesidad de velar por la coherencia de las políticas de seguridad, desarrollo, comercio, y las políticas humanitarias y militares.

Estos criterios de evaluación han de considerarse «conceptos» que es preciso tener en cuenta en la evaluación. No obstante, no se han definido con suficiente detalle como para aplicarlos de forma sistemática y transparente en la formulación de juicios evaluativos sobre los programas o las políticas. En el marco de cada uno de los criterios «genéricos», deberían definirse y acordarse con los principales interesados otros criterios más específicos, tales como normas de referencia o estándares⁸, adecuados a cada tipo de intervención y a su contexto.

Por ejemplo, en el caso de una evaluación global sobre la aplicación en la programación de UNICEF del enfoque basado en los derechos humanos, los evaluadores identificaron una serie de planteamientos diferentes sobre lo que este representa. Llevaron a cabo consultas y sesiones de formación de consenso en las que se otorgó prioridad a cinco principios clave (enumerados anteriormente) y posteriormente

⁴ Véase Fondo de las Naciones Unidas para la Infancia, Trabajar por un futuro de igualdad, Política de UNICEF sobre igualdad de género y empoderamiento de las mujeres y las niñas, UNICEF, Nueva York, 2010.

⁵ Véase Fondo de las Naciones Unidas para la Infancia, *Evaluación mundial de la aplicación del enfoque basado en los derechos humanos a la programación de UNICEF*, informe final - volumen I, UNICEF, Nueva York, 2012. (Nota: se consideró que un sexto principio, el de indivisibilidad o interdependencia, era inherente al principio de normatividad).

⁶ Parafraseados del capítulo «Linking evaluation criteria to evaluation questions: Additional criteria for humanitarian assistance programmes», disponible en *UNICEF Monitoring and Evaluation Training Resource*. El documento todavía se está revisando, de ahí que no se indiquen las páginas y que solo se facilite un fragmento.

⁷ Comité de Asistencia para el Desarrollo de la Organización de Cooperación y Desarrollo Económicos, *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, OCDE-CAD, París, 1999.

⁸ Una norma de referencia o índice es un conjunto de indicadores conexos que facilitan comparaciones significativas, precisas y sistemáticas sobre el desempeño; un estándar o rúbrica es un conjunto de normas de referencia, índices o indicadores que brinda información significativa desde el punto de vista del desempeño (véase la Síntesis n.º 11 (Desarrollo y selección de medidas de bienestar infantil) y la Síntesis n.º 4 (Razonamiento evaluativo)).

definieron los indicadores principales que deben aplicarse en la evaluación de los programas específicos de UNICEF.

Puntos principales

1. Los criterios de evaluación especifican los valores que se emplearán en la evaluación de una intervención (es decir, de un programa o una política).
2. Las evaluaciones de UNICEF recurren a diversos criterios de evaluación. No todos se aplican en todas las evaluaciones, dado que algunos solo resultan adecuados para determinadas intervenciones o tipos de evaluación. Las evaluaciones de impacto deben prestar atención a:
 - la pertinencia, la eficacia, la eficiencia, el impacto y la sostenibilidad;
 - la equidad, la igualdad de género y los derechos humanos; y
 - la cobertura, la coordinación, la protección y la coherencia (en especial con respecto a la asistencia humanitaria).
3. Los criterios de evaluación «genéricos» deben dotarse de mayor detalle en los términos de referencia de una evaluación o en las etapas iniciales del proceso, para que puedan aplicarse de forma sistemática y transparente en la formulación de juicios evaluativos sobre la intervención.

2. ¿CUÁNDO PROCEDE EMPLEAR LOS CRITERIOS DE EVALUACIÓN?

Todas las evaluaciones de impacto que se llevan a cabo en el contexto de la labor de UNICEF deberían tener en cuenta los criterios del CAD de la OCDE y determinar cuáles resultan pertinentes en cada caso. Esos criterios reflejan los principios fundamentales para la evaluación de la asistencia para el desarrollo⁹ y han sido adoptados como norma de buenas prácticas por la mayoría de los organismos que trabajan en este ámbito. Los criterios de equidad, igualdad de género y derechos humanos son también de aplicación obligada en las evaluaciones de impacto (y de otro tipo) de UNICEF.

En función del tipo de intervención (por ejemplo, un caso de asistencia humanitaria como el mencionado anteriormente) o de evaluación (por ejemplo, una evaluación de procesos), cabrá aplicar otros criterios o centrarse en un conjunto determinado. Por ejemplo, el criterio de impacto del CAD de la OCDE no resulta pertinente en una evaluación de procesos, dado que esta examina cómo se ejecuta una intervención (entre otros aspectos, cómo se prestan los servicios, si los clientes se sienten satisfechos con ellos y cuáles son las prácticas de gestión), en lugar de si esta ha propiciado los resultados previstos.

3. ¿CÓMO APLICAR LOS CRITERIOS DE EVALUACIÓN?

Vincular los criterios con las preguntas de evaluación

Los términos de referencia de la evaluación deben especificar su objetivo general y alcance, y los criterios genéricos que se tendrán en cuenta. Asimismo, deben establecer las [preguntas clave de evaluación](#) (PCE), si bien estas pueden seguir desarrollándose o ajustándose como primer paso del proceso (véase la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)).

⁹ Organización de Cooperación y Desarrollo Económicos, Comité de Asistencia para el Desarrollo, *Principles for Evaluation of Development Assistance*, OCDE-CAD, París, 1991. Véase <http://www.oecd.org/dac/evaluation/50584880.pdf>.

A fin de mantener el énfasis de la evaluación de impacto, únicamente debe especificarse un conjunto limitado de PCE o preguntas de evaluación de alto nivel (macronivel). Por ejemplo:

- PCE1. ¿Qué nivel de calidad tuvo el diseño o contenido de la intervención? Pertinencia, equidad, igualdad de género, enfoque de programación basado en los derechos humanos.
- PCE2. ¿En qué medida se ejecutó y adaptó correctamente la intervención según las necesidades? Eficacia y eficiencia.
- PCE3. ¿Obtuvo la intervención los resultados previstos a corto, medio y largo plazo? De ser así, ¿en favor de quién, en qué medida y en qué circunstancias? Eficacia, impacto, equidad, igualdad de género.
- PCE4. ¿Qué impactos imprevistos (positivos y negativos) generó la intervención? ¿Cómo se produjeron? Eficacia, impacto, equidad, igualdad de género, enfoque de programación basado en los derechos humanos.
- PCE5. ¿Qué obstáculos y factores facilitadores marcaron la diferencia entre el éxito o el fracaso en la ejecución y los resultados de la intervención? Pertinencia, equidad, igualdad de género, enfoque de programación basado en los derechos humanos.
- PCE6. ¿En qué medida se lograron resultados valiosos para los proveedores de servicios, los clientes, la comunidad y las organizaciones participantes? Pertinencia, equidad, igualdad de género, enfoque de programación basado en los derechos humanos.
- PCE7. ¿En qué medida hizo la intervención un uso óptimo de los recursos disponibles para lograr los resultados más valiosos posibles en favor de los participantes y la comunidad? Eficiencia.
- PCE8. ¿Es probable que se sostenga algún resultado positivo? ¿En qué circunstancias? Sostenibilidad, equidad, igualdad de género, enfoque de programación basado en los derechos humanos.

A continuación deben formularse un conjunto de preguntas de evaluación más pormenorizadas (de nivel medio e inferior) que aborden en detalle cada uno de los criterios de evaluación. Todas esas preguntas deben vincularse expresamente con los criterios de evaluación, para cubrirlos en su totalidad. Los cuadros 1 a 5 recogen ejemplos de preguntas de nivel medio (mesonivel) relativas a los criterios del CAD de la OCDE, que posteriormente se desglosan en preguntas de nivel inferior (micronivel). Al hablar de niveles alto/macro, medio/meso e inferior/micro no se pretende establecer una jerarquía en cuanto a su importancia; reflejan sencillamente un grado creciente de concreción u operativización.

Cuadro 1. Ejemplos de preguntas de evaluación de mesonivel y micronivel para determinar la «pertinencia»

¿Fue la intervención acorde con los compromisos y el plan estratégico del país encaminados a mejorar el bienestar de los niños?

- ¿Fueron los objetivos de la intervención pertinentes a las necesidades y prioridades nacionales identificadas?

¿Fue el diseño de la intervención correcto desde el punto de vista técnico?

- ¿Fueron los objetivos y el diseño de la intervención adecuados al contexto y a las necesidades de sus beneficiarios?
- ¿Siguió la intervención una lógica clara y disponía de una base de pruebas pertinente en la zona que vinculase sus actividades y operaciones con las realidades a que hacen frente sus

beneficiarios? ¿Se justificaron de manera clara las actividades seleccionadas en el marco de la intervención?

- ¿Fueron las actividades y los productos de la intervención conformes con la meta general y el logro de sus objetivos? ¿Fueron asimismo acordes con los resultados e impactos previstos?
- ¿Reflejaron adecuadamente los fondos asignados el nivel de las necesidades y el entorno operacional?

¿Fomentó la intervención la igualdad de género y otras normas sobre derechos humanos?

- ¿En qué medida se incorporaron a la planificación de la intervención las metas y los procesos de equidad, igualdad de género y enfoque de programación basado en los derechos humanos?
- ¿Se previó en el diseño de la intervención la participación igualitaria de todos los grupos correspondientes (por ejemplo, hombres y mujeres, niños y niñas)?
- ¿Promovió la intervención un acceso más igualitario de hombres y mujeres, niños y niñas a los beneficios de la actividad y, en un sentido más amplio, a los recursos, los servicios y las aptitudes?

¿Se ajustó la intervención en su fase de ejecución para que fuera acorde con las nuevas prioridades y necesidades, y velara por la aplicación de las mejores prácticas?

- ¿En qué medida tuvieron cabida en la intervención actividades y procesos o mecanismos que favorecieran un diseño y una ejecución eficaces del programa y recabaran las impresiones de los interesados?

Cuadro 2. Ejemplos de preguntas de evaluación de mesonivel y micronivel para determinar la «eficacia»

¿Se ejecutó la intervención de acuerdo con el plan previsto?

- ¿Se ejecutó la intervención de acuerdo con el plan previsto? De no ser así, ¿por qué? ¿Qué medidas se tomaron al respecto?
- ¿En qué medida facilitaron u obstaculizaron la ejecución de la intervención los factores contextuales?

¿Se tomaron medidas correctivas en el momento oportuno en caso necesario?

- ¿Se identificó la necesidad de apoyo complementario o se facilitó para solventar problemas en la ejecución? ¿Qué forma adoptó tal apoyo, quién lo facilitó y con qué efecto?
- ¿Qué riesgos amenazaban la consecución de los objetivos de la intervención? ¿Se gestionaron esos riesgos de forma adecuada?

¿Se lograron los resultados previstos?

- ¿En qué medida se alcanzaron los objetivos de la intervención? ¿Cuáles fueron los factores que más favorecieron o impidieron el logro de los objetivos?

Cuadro 3. Ejemplos de preguntas de evaluación de mesonivel y micronivel para determinar la «eficiencia»

¿Usó la intervención los recursos disponibles de la manera más rentable para lograr sus objetivos?

- ¿Se alcanzaron los objetivos a tiempo?
- ¿Dispuso la intervención de una dotación de personal suficiente y apropiada? Al ejecutar la intervención, ¿se utilizaron el tiempo y los recursos con eficacia para lograr resultados?
- ¿En qué medida se persiguió y logró una coordinación y colaboración eficaz con las intervenciones en curso y los asociados?
- ¿Se diseñó o corrigió la intervención a lo largo de la fase de ejecución en aras de la mejor relación calidad-precio?
- ¿Fue la intervención eficiente en igual o mayor medida que otros planteamientos adecuados en el mismo contexto?
- ¿Existen otros modos viables de ejecutar la intervención que resultarían más rentables?

Cuadro 4. Ejemplos de preguntas de evaluación de mesonivel y micronivel para determinar el «impacto»

¿Se amplió lo suficiente la escala de la intervención para lograr los impactos previstos (por ejemplo, en los indicadores clave de salud, las desigualdades)?

- ¿Se obtuvieron pruebas de un cambio, ya sea positivo o negativo? De ser así, ¿qué factores contribuyeron a él? De no ser así, ¿por qué?
- ¿Hubo cambios imprevistos —positivos o negativos— en la vida de los beneficiarios y su entorno? ¿Cuáles? ¿Se relacionaron directa o indirectamente con el programa o se debieron a factores externos?

¿Se lograron los resultados de manera acorde con la equidad, la igualdad de género y otros derechos humanos?

- ¿En qué grado consiguió la intervención que participaran tanto las mujeres como los hombres, los titulares de derechos y sus garantes, especialmente los más vulnerables?
- ¿En qué medida se beneficiaron —de formas diversas— distintos grupos, entre ellos los niños y otros grupos vulnerables, de la intervención?
- ¿En qué medida aumentó la capacidad de distintos grupos, entre ellos los niños y otros grupos vulnerables, para relacionarse de manera diferente con otros participantes en la intervención?
- ¿Cómo se distribuyeron entre los grupos de participantes los costos y las cargas de la participación en la intervención? ¿Y los beneficios comparativos, si surgió alguno de su participación?

¿Influyó el contexto operacional en la intervención, o viceversa?

- ¿Qué función desempeñó el contexto local y nacional en favor o en contra del cambio?
- ¿Tuvo impacto la intervención en el entorno operacional más amplio donde se trabaja en aras del bienestar de los niños del país? De ser así, ¿qué forma adoptó?

Cuadro 5. Ejemplos de preguntas de evaluación de mesonivel y micronivel para determinar la «sostenibilidad»

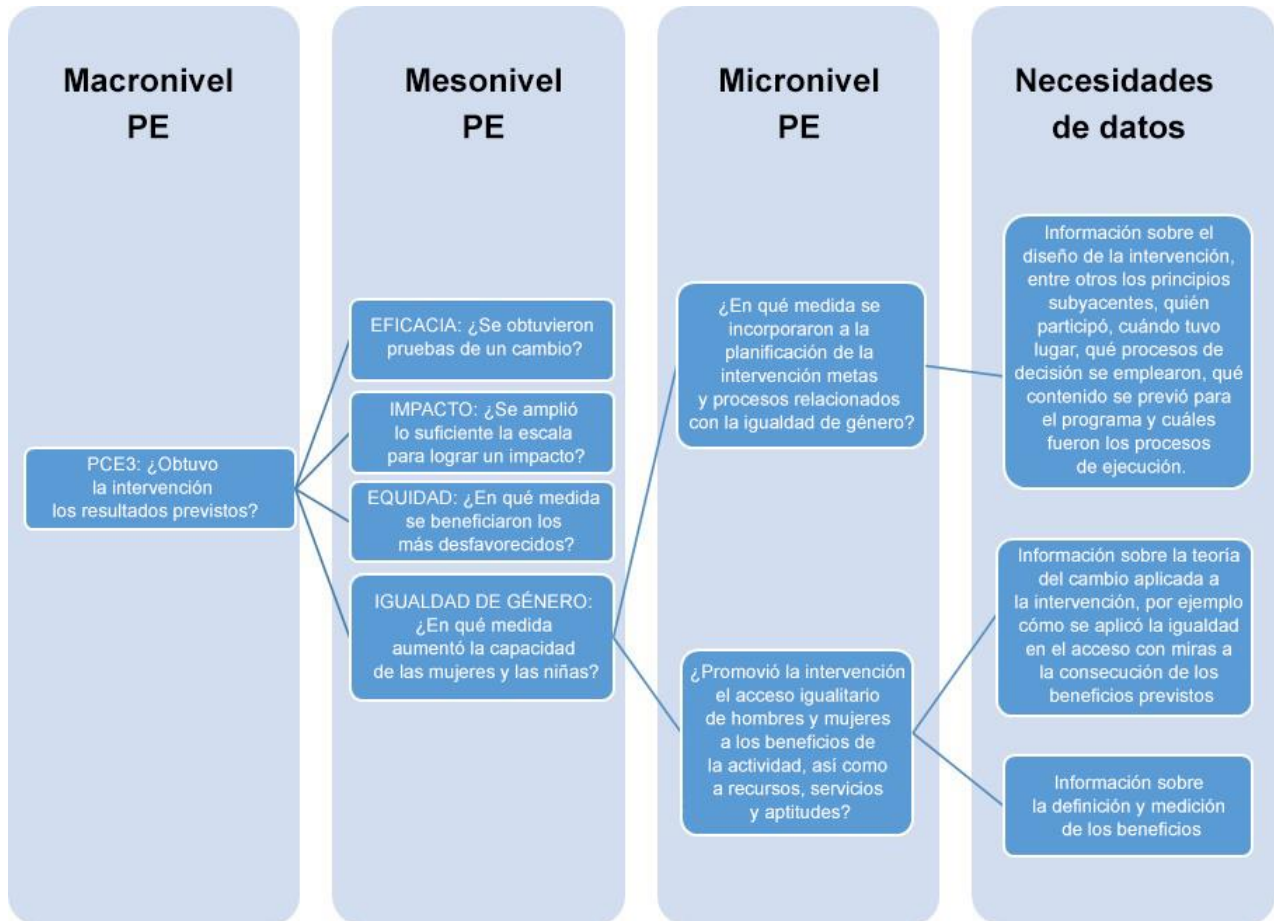
¿Se lograron los resultados de forma sostenible?

- ¿En qué medida pueden las actividades y los beneficios de la intervención tener continuidad una vez que se interrumpa la financiación externa?
- ¿Existen ámbitos de la intervención claramente insostenibles? ¿Qué lecciones se pueden extraer de esos ámbitos?
- ¿En qué medida han asumido los beneficiarios de la intervención o los asociados nacionales el proyecto como propio y están dotados de la capacidad y los recursos necesarios para mantener los resultados de la actividad una vez que se interrumpa la financiación externa?
- ¿Contribuyó la intervención a un aumento de la capacidad de las organizaciones locales para seguir brindando servicios de calidad dirigidos a los niños? De ser así, ¿qué forma adoptó?
- ¿Qué factores facilitaron o impidieron en mayor medida que la intervención fuera sostenible?

En respuesta a los términos de referencia de la evaluación debe elaborarse un plan de evaluación en el que se describan las preguntas de evaluación clave y pormenorizadas en relación con los criterios genéricos y específicos (o cómo se determinarán estos últimos en el proceso de evaluación); los datos necesarios para responder a cada una de las preguntas y el diseño de la investigación conexo (véase la Síntesis n.º 6 (Sinopsis: Estrategias de atribución causal)); así como los métodos de recopilación o cotejo y análisis de datos (véase la Síntesis n.º 10 (Sinopsis: Métodos de recolección y análisis de datos en la evaluación de impacto)). El plan de evaluación debe establecer también las responsabilidades (es decir, quién se encargará de cada aspecto), los plazos y los entregables que servirán a los comisionados o directores de la evaluación para efectuar un seguimiento del proceso.

El gráfico 1 resume la relación entre los criterios y las preguntas de evaluación (a los niveles macro, meso y micro) y las necesidades de datos en el ejemplo de la PCE3 (véase más arriba): ¿Produjo la intervención los resultados previstos a corto, medio y largo plazo? De ser así, ¿en favor de quién, en qué medida y en qué circunstancias?

Gráfico 1. Ejemplo ilustrativo de cómo se vincula una pregunta de evaluación con el criterio correspondiente y las necesidades de datos específicas



Nota: PE = pregunta de evaluación.

Uso de normas de referencia o estándares específicos para juzgar el desempeño de las intervenciones

Las preguntas de evaluación adecuadas no se interesan únicamente por cuáles fueron los resultados (es decir, preguntas descriptivas sobre cómo son las cosas y qué ha sucedido) sino también por *el grado de idoneidad* de los resultados (es decir, preguntas valorativas sobre la conclusión general en cuanto a si el programa o la política lograron de forma adecuada aquello que pretendían).

Es posible que los términos de referencia de la evaluación incluyan normas de referencia o estándares específicos en los que deba basarse la valoración de la intervención, o, si estos no existen todavía, indicaciones sobre cómo han de identificarse y acordarse durante el proceso de evaluación. Por ejemplo, si definimos la eficacia como la medida en que se logran los objetivos, y uno de los objetivos es «que los niños pasen más tiempo en la escuela», es probable que los criterios específicos a partir de los que se juzgará el valor de la intervención hagan hincapié en la retención (es decir, proseguir los estudios hasta los últimos cursos) y la asistencia (por ejemplo, que todos los niños asistan como mínimo al 90% de las clases, o que el 90% de los alumnos asistan a todas las clases).

Existen normas de referencia o estándares para determinados sectores o contextos relacionados con la infancia. Generalmente se basan en pruebas y se desarrollan en consulta con los interesados pertinentes. Su propósito fundamental es facilitar metas concretas (más realistas que ambiciosas) con las que comparar los progresos (en los ámbitos local, nacional y mundial) y evaluar los logros. Si se definen y ejecutan correctamente, sirven para comparar y aprender del desempeño de distintos programas y políticas, situaciones y zonas geográficas concretas. Por ejemplo, se definieron normas de referencia o estándares para:

- [Metas de la OMS relativas a la cobertura vacunal](#): La justificación es que una cobertura mínima es fundamental para prevenir con eficacia los brotes de enfermedades.
- [Sistemas nacionales para la protección de la infancia de UNICEF](#) (véase el cuadro 6): La justificación es que los sistemas nacionales de protección de la infancia gozan de reconocimiento como el medio más eficaz y sostenible para proteger a los niños de toda forma de malos tratos. La inexistencia de este tipo de sistemas impide que la cobertura sea plena, de manera que cada vulneración se trata como un problema independiente e inconexo¹⁰.

Los criterios de evaluación no solo resultan útiles al valorar las intervenciones, sino que también deben reflejarse de manera expresa en la [teoría del cambio](#) (véase la Síntesis n.º 2 (La teoría del cambio)) y aplicarse en el contenido de la intervención y en su modo de ejecución. El cuadro 6 muestra un ejemplo de UNICEF sobre las normas de referencia y los estándares que se proponen para las evaluaciones en el sector de la protección de la infancia.

Cuadro 6. Normas de referencia y estándares empleados en la prestación y la evaluación de los servicios de protección de la infancia

Meta

Con vistas a brindar apoyo al personal que trabaja en la prestación de servicios de protección de la infancia, es esencial que los países pongan en marcha sistemas integrales que den cabida a los requisitos legales correspondientes, a un proceso continuo de formación relacionada y a un sistema de inspección periódica del nivel de las aptitudes.

Normas de referencia del programa

La Oficina Regional para Asia Oriental y el Pacífico de UNICEF propuso las siguientes **normas de referencia** de cara a un sistema de formación profesional y perfeccionamiento continuo:

- Una titulación universitaria en trabajo social, cuyo plan de estudios incluya cursos sobre los servicios sociales, cuestiones relacionadas con el desarrollo, temas de protección y prevención, e intervenciones terapéuticas.
- Una cualificación profesional en trabajo social o desarrollo infantil, con un plan de estudios aprobado por las autoridades competentes.
- Un sistema de acreditación de las aptitudes para el trabajo social, basado en pruebas de competencia en el marco de los programas de capacitación pertinentes.
- Capacitación de los trabajadores del ámbito de la educación (entre otros, los profesores), los trabajadores sanitarios y otros profesionales que trabajan con niños sobre cómo hacer frente al abuso, la violencia y la explotación.

¹⁰ Fondo de las Naciones Unidas para la Infancia, *Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific*, Strengthening Child Protection Systems No. 3, UNICEF EAPRO, Bangkok, 2012, prólogo, p. vii. Véase http://www.unicef.org/eapro/Measuring_and_monitoring.pdf.

- Un programa de capacitación específico para los funcionarios ministeriales con una función principal en los asuntos de interior o justicia que afectan a los niños y la justicia.

Juicios evaluativos

A continuación se especificaron una serie de **estándares** para valorar la situación del sistema de capacitación profesional de los países:

- Un sistema educativo de nivel A incluye **todos** los elementos de los criterios de referencia.
- Un sistema educativo de nivel B incluye **al menos tres** elementos de los criterios de referencia.
- Un sistema educativo de nivel C incluye **uno o dos** elementos de los criterios de referencia.
- Un sistema educativo de nivel D no incluye **ningún** elemento de los criterios de referencia.

Los estándares anteriores pueden emplearse en la valoración de un programa nacional de formación profesional; entre otros, como fundamento de un sistema de calificación. Por ejemplo:

- **Excelente:** Modelo inequívoco de desempeño ejemplar o mejor práctica en su ámbito; carece de puntos débiles; cumple el nivel A.
- **Notable:** Desempeño muy bueno o excelente en prácticamente todos los aspectos; sólido en general, aunque no ejemplar; sin puntos débiles con verdadera repercusión; cumple los requisitos del nivel B y, además, ha superado todos los requisitos de calidad indicados con respecto a la formación.
- **Inaceptable:** Deficiencias importantes en todos los ámbitos del programa; cumple los requisitos del nivel C o D.

Fuente: Fondo de las Naciones Unidas para la Infancia, *Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific*, Strengthening Child Protection Systems No. 3, UNICEF EAPRO, Bangkok, 2012, prólogo, p. vii. Véase http://www.unicef.org/eapro/Measuring_and_monitoring.pdf.

La reflexión sobre cómo determinar —y por tanto valorar— el éxito de un programa o una política debería tener lugar por adelantado. De ese modo, será posible dirigir la recopilación de datos hacia el conjunto de pruebas necesario para valorar el desempeño de forma adecuada. También resulta más transparente y objetivo definir el «éxito» antes de que se reúnan las pruebas. Implicar a los interesados pertinentes en la definición de los criterios de éxito de una intervención determinada en su contexto concreto propicia una mayor aceptación de las conclusiones de la evaluación (al reducir los malentendidos y desacuerdos relativos a su interpretación posterior) y promueve un compromiso con vistas a la adopción de las medidas necesarias en relación con el programa. La Síntesis n.º 4 (Razonamiento evaluativo) establece un enfoque gradual a la hora de decidir 1) qué dimensiones de la intervención deben funcionar (es decir, fijar criterios específicos de valor relativos a la «equidad») y 2) qué grado de desempeño debe tener la intervención en esas dimensiones (es decir, elaborar un estándar o [rúbrica](#), o bien emplear uno que ya exista).

4. CUESTIONES ÉTICAS Y LIMITACIONES PRÁCTICAS

Cuestiones éticas

Todas las evaluaciones de UNICEF deben aplicar las *Ethical Guidelines for Evaluation* (Directrices éticas de evaluación)¹¹ establecidas por el Grupo de Evaluación de las Naciones Unidas (UNEG). Estas directrices establecen tanto un código de conducta para evaluadores como obligaciones éticas para los participantes en la evaluación (véase asimismo la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)). Estas incluyen: respeto de la dignidad y la diversidad; derechos; confidencialidad; y prevención del daño. Las obligaciones anteriores deben cumplirse explícitamente en el diseño y durante todo el proceso de evaluación. Su intención es garantizar el respeto y la protección de los derechos y el bienestar de los seres humanos y las comunidades a las que pertenecen, de conformidad con la Declaración Universal de Derechos Humanos y otros convenios de derechos humanos.

En octubre de 2013, el Centro de Investigaciones Innocenti de UNICEF puso en marcha [Ethical Research Involving Children](#), un recurso en línea¹² que reúne planteamientos expertos sobre importantes cuestiones éticas relacionadas con la infancia y cómo podrían abordarse en distintos contextos de investigación (y evaluación).

Además de las cuestiones éticas generales mencionadas anteriormente, existen una serie de cuestiones particulares relacionadas con la elección de los criterios de evaluación, tales como:

- ¿Quién debería participar en la definición del éxito? Ha de decidirse quién tiene la última palabra cuando surjan controversias entre distintas prioridades y puntos de vista o aparezcan pruebas contradictorias.
- ¿Cómo se define el éxito? Si se define el éxito en términos de «efectos medios» o «indicadores sociales agregados», es posible que se deje de lado la experiencia de ciertos subgrupos (por ejemplo, los niños o las niñas, las personas más vulnerables o las más pobres).
- ¿Cómo se lograron los resultados previstos? Por ejemplo, si se pretende incrementar la cobertura vacunal, ¿es poco ético que las autoridades obliguen a la población a participar a fin de hacer frente a un brote?

Limitaciones prácticas

Entre las limitaciones prácticas habituales en relación con los criterios de evaluación cabe destacar:

- Las dificultades de la transición desde una redacción genérica de los criterios de evaluación (por ejemplo, el criterio de «sostenibilidad» del CAD de la OCDE) a criterios más específicos que tengan en cuenta las opiniones de diversos interesados sobre qué dimensiones son importantes y qué umbrales deberían superar las intervenciones.
- La discordancia entre el calendario de la evaluación de impacto y el momento en que los impactos pertinentes o las pruebas de la sostenibilidad serán visibles. Las mediciones indirectas (es decir,

¹¹ Grupo de Evaluación de las Naciones Unidas, *Ethical Guidelines for Evaluation*, UNEG, Nueva York, 2007.

¹² Los nuevos recursos son la International Charter for Ethical Research Involving Children (Carta internacional para la ética en la investigación con niños); un sitio web, www.childethics.com, diseñado específicamente con objeto de ofrecer un abundante depósito de información con base empírica, recursos y enlaces a artículos de revistas dirigidos a orientar y mejorar la investigación con niños y proporcionar una plataforma para seguir manteniendo una reflexión y un diálogo fundamentales; y un compendio de cuestiones y desafíos éticos, incluida una colección de más de 20 casos de estudio, así como preguntas estructuradas para orientar la ética en la investigación con niños (la página web llamada «[Getting Started](#)»).

aquello que es posible medir, en lugar de lo que debería medirse idealmente) pueden complicar todavía más las definiciones y las negociaciones en torno a los criterios de éxito.

- Con frecuencia disponemos de información escasa sobre la calidad de determinados componentes o servicios que se ejecutan en el marco de un programa a gran escala. Es difícil reconstruir o comprender esa información retrospectivamente, y por tanto difícil capturarla de manera adecuada al especificar los criterios de evaluación.

Si los recursos y el marco temporal de la evaluación lo permiten, algunas de esas limitaciones pueden corregirse (por ejemplo ampliando el periodo o el alcance de la evaluación), pero es probable que las conclusiones tengan que extraerse sobre la base de datos imperfectos. Por lo tanto, es importante ser transparente respecto a las limitaciones y describir cómo pueden haber afectado a las constataciones, las conclusiones y las recomendaciones generales de la evaluación.

5. ¿QUÉ OTROS MÉTODOS FUNCIONAN BIEN CON ESTE?

Todas las evaluaciones tienen la misión de plantear y responder a unas preguntas de evaluación —no deben limitarse a medir y describir resultados—. Así pues, todas sus conclusiones han de interpretarse en un marco evaluativo que no solo indique cuáles son los resultados, sino también su grado de idoneidad. Para ello es preciso llevar a cabo un razonamiento evaluativo, obligatorio en todas las evaluaciones con independencia de sus métodos o enfoques (véase la Síntesis n.º 4 (Razonamiento evaluativo)).

Dado que, en el contexto del trabajo de UNICEF, numerosas evaluaciones de impacto se basarán en las fuentes de datos existentes, también es importante incluir una evaluación basada en los estándares¹³ sobre la calidad del sistema de seguimiento y evaluación, vinculada al programa y al entorno de políticas más amplio. El uso de datos de calidad desconocida puede poner en entredicho la validez de las constataciones de una evaluación y dar pie a conclusiones engañosas. Los evaluadores deben utilizar o recopilar datos de la mayor calidad y minimizar, en la medida de lo posible, las consecuencias de cualquier duda sobre la calidad de los datos, que señalarán expresamente en el informe de evaluación.

6. PRESENTACIÓN DE LOS RESULTADOS Y ANÁLISIS

La estructura del informe de evaluación debe reflejar el propósito y las PCE de la evaluación (véase también la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)).

En el primer caso, han de resumirse las pruebas que responden a las preguntas pormenorizadas relacionadas con los criterios de pertinencia, eficacia, eficiencia, impacto y sostenibilidad del CAD de la OCDE, si bien con un grado de detalle suficiente que justifique las conclusiones y recomendaciones. Exponer las pruebas de ese modo ayudará también al [equipo de gestión de la evaluación](#) a comprobar que se cumplen plenamente los criterios estipulados en los términos de referencia.

Deben emplearse rúbricas evaluativas específicas para «interpretar» las pruebas y determinar qué consideraciones son de importancia o urgencia fundamental. A continuación, deben sintetizarse las pruebas sobre múltiples dimensiones y generar respuestas a las preguntas de evaluación de alto nivel. En la Síntesis n.º 4 (Razonamiento evaluativo) se facilitan recomendaciones más específicas sobre cómo

¹³ Existen herramientas disponibles sobre normas mundiales acordadas para la evaluación de los sistemas de seguimiento y evaluación (incluidos los planos nacional, subnacional y de prestación de servicios). Por ejemplo, Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA, *12 Components Monitoring and Evaluation System Strengthening Tool*, ONUSIDA, Ginebra, 2009. Véase http://www.unaids.org/en/media/unaids/contentassets/documents/document/2010/2_MERG_Strengthening_Tool_12_Component_s_ME_System.pdf.

redactar un informe con síntesis evaluativas sólidas (por ejemplo, usando un lenguaje evaluativo expreso al presentar las conclusiones, en lugar de un lenguaje neutro que se limite a describirlas).

7. EJEMPLOS DE BUENAS PRÁCTICAS

En el cuadro 7 se resumen algunos ejemplos concretos de buenas prácticas.

Cuadro 7. Ejemplo de una evaluación de impacto en curso relativa a los programas conjuntos sobre género del sistema de las Naciones Unidas

Buenas prácticas en los términos de referencia

Se incluyeron siete PCE que se vincularon directamente con las prioridades estratégicas y los problemas nuevos identificados en el análisis de la cartera del programa, por ejemplo:

- ¿En qué medida se conceptualizaron, planificaron y diseñaron conjuntamente los programas sobre género para dar respuesta a los compromisos internacionales, regionales y nacionales sobre la igualdad de género y el empoderamiento de la mujer; promover la coherencia y aprovechar las ventajas comparativas de los organismos de las Naciones Unidas participantes; e integrar un enfoque de programación basado en los derechos humanos?
- ¿Cuáles son las contribuciones principales y el valor añadido en relación con los resultados a corto y largo plazo, previstos e imprevistos, positivos y negativos, alcanzados hasta la fecha en el plano nacional en el ámbito de la igualdad de género y el empoderamiento de la mujer por los programas conjuntos sobre género?
- ¿Cómo influyeron el contexto nacional y las circunstancias específicas (por ejemplo, situación de conflicto o posconflicto, fragilidad estatal, situación de la iniciativa «Unidos en la acción») en el logro de los resultados y la eficacia operacional del programa conjunto sobre género?

Los criterios para la evaluación de los programas conjuntos sobre género también se esbozaron y definieron en el contexto del programa (que no se ciñó por tanto a la descripción genérica de los criterios del CAD de la OCDE). Por ejemplo:

- Pertinencia y coherencia de los procesos de planificación, diseño y ejecución del programa conjunto sobre género con respecto a los compromisos, las políticas y las prioridades regionales, nacionales e internacionales; con los principios de eficacia de la ayuda; con los mandatos de las Naciones Unidas y el Marco de Asistencia de las Naciones Unidas para el Desarrollo; y con las políticas, los mandatos y las ventajas comparativas de cada organismo, en cuanto a su conformidad y armonización con las necesidades de los países en materia de igualdad de género y empoderamiento de la mujer.
- Eficacia e impacto de los programas conjuntos sobre género en el logro de los objetivos fijados en favor de la igualdad de género y el empoderamiento de la mujer, y todo efecto a largo plazo, previsto o imprevisto, en el plano nacional y de cara a la transversalización de la cuestión de género en el sistema de las Naciones Unidas, incluido el uso de planteamientos innovadores.
- Sostenibilidad de los resultados de los programas conjuntos sobre género en vista de la asunción nacional del proyecto como propio, la forja de alianzas eficaces y el fortalecimiento de la capacidad nacional a través de los procesos de tales programas.

Se introdujeron otros criterios en los procesos de los programas conjuntos sobre género; por ejemplo, la participación y la inclusión de los garantes y titulares de derechos —concretamente, los más marginados—.

En el periodo de planificación también se añadieron otros criterios antes de concluir el alcance final de la evaluación.

Aparte de las PCE, en un anexo a los términos de referencia se incluyó una «matriz de preguntas de evaluación» con preguntas más detalladas.

Fuente: Véanse los términos de referencia relacionados con las evaluaciones y los informes temáticos vinculados con el Fondo de Población de las Naciones Unidas, «Joint Evaluation of Joint Gender Programmes in the UN System», UNPFA, <http://www.unfpa.org/admin-resource/joint-evaluation-joint-gender-programmes-un-system>.

8. EJEMPLOS DE POSIBLES PROBLEMAS

Algunas deficiencias habituales relacionadas con los criterios de evaluación se identificaron, por ejemplo, en el examen de la calidad de todos los informes de evaluación de 2012¹⁴.

- «No se especificaron las preguntas y criterios subyacentes en que se fundamenta la evaluación».
- «La integración en las evaluaciones del enfoque de programación basado en los derechos humanos, la igualdad de género y la equidad sigue siendo por lo general deficiente (es decir, el porcentaje de informes que integraron correctamente esos elementos no superó el 50%; la cuestión de género se trató en el 46% de los informes, el 44% prestó atención al enfoque de programación basado en los derechos humanos, y el 41% incorporó un énfasis especial en la equidad)».

Se señaló asimismo: «Los informes de evaluación reflejan por lo general aquello que se pidió que los evaluadores considerasen, incluidos los métodos sugeridos, los criterios de evaluación, las fuentes de datos y las preguntas destacadas en los [términos de referencia]. Aunque las cuestiones de género, derechos humanos y equidad son fundamentales para los valores del organismo, su tratamiento en los informes de evaluación suele estar supeditado a los [términos de referencia]».

Por tanto, es importante que a lo largo del proceso los evaluadores reciban directrices, términos de referencia y observaciones coherentes y explícitos.

¹⁴ En el examen se revisaron 79 informes. Las citas proceden del resumen (p. iii) de Universal Management Group, «GEROS – Global Meta-Evaluation Report 2012», informe final, Universal Management Group, 2013. Véase http://www.unicef.org/evaldatabase/files/2013_GEROS_MetaEvaluation_finalReport.pdf.

9. LECTURAS Y ENLACES CLAVE

BetterEvaluation, «Define ethical and quality evaluation standards», página web, BetterEvaluation, http://betterevaluation.org/plan/manage_evaluation/ethical_evaluation.

BetterEvaluation, «Establish Decision Making Processes», página web, BetterEvaluation, http://betterevaluation.org/plan/manage/who_controls.

BetterEvaluation, «Evaluation Plan», página web, BetterEvaluation, http://betterevaluation.org/evaluation-options/evaluation_planning_templates.

BetterEvaluation, «Specify the Key Evaluation Questions», página web, BetterEvaluation, http://betterevaluation.org/plan/engage_frame/decide_evaluation_questions.

Centro de Investigaciones Innocenti de UNICEF, Childwatch International Research Network, Centre for Children and Young People de la Universidad Southern Cross (Australia) y Children's Issues Centre de la Universidad de Otago (Nueva Zelanda), «Ethical Research Involving Children», sitio web, 2013, www.childethics.com.

Grupo de Evaluación de las Naciones Unidas, *Ethical Guidelines for Evaluation*, UNEG, Nueva York, 2007. Véase <http://www.unevaluation.org/ethicalguidelines>.

Organización de Cooperación y Desarrollo Económicos, «Evaluation of development programmes, DAC Criteria for Evaluating Development Assistance», página web, OCDE, <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

Organización de Cooperación y Desarrollo Económicos, Comité de Asistencia para el Desarrollo, *Principles for Evaluation of Development Assistance*, OCDE-CAD, París, 1991. Véase <http://www.oecd.org/dac/evaluation/50584880.pdf>.

GLOSARIO

<u>Equipo de gestión de la evaluación (EGA)</u>	<i>Equipo a cargo de gestionar una evaluación. En UNICEF, es responsable de tareas como seleccionar y supervisar el trabajo de los evaluadores (normalmente externos); revisar los términos de referencia, el informe inicial y el proyecto de informe final; brindar asesoramiento técnico a los evaluadores; gestionar los aspectos operacionales de la evaluación (presupuesto, visitas sobre el terreno, contacto con los homólogos pertinentes de UNICEF), y garantía de la calidad. En una evaluación conjunta, incluiría al personal de UNICEF y a representantes de cada organización asociada. Términos conexos: comité directivo de evaluación.</i>
<u>Preguntas clave de evaluación (PCE)</u>	<i>Preguntas de evaluación de alto nivel (macronivel) sobre el desempeño general, a las que la evaluación debe tratar de dar respuesta. Las PCE se derivan del propósito de la evaluación.</i>
<u>Rúbrica</u>	<i>Herramienta de evaluación que utiliza un conjunto de criterios y una guía de calificación/puntuación predeterminada por los evaluadores. Las rúbricas pueden emplearse para evaluar presentaciones, proyectos y carteras, entre otros.</i>
<u>Teoría del cambio</u>	<i>Explica cómo se entiende que las actividades produzcan una serie de resultados que contribuyen a lograr los impactos finales previstos. Puede elaborarse para cualquier nivel de intervención: un acontecimiento, un proyecto, un programa, una política, una estrategia o una organización.</i>
<u>Términos de referencia</u>	<i>Descripción de los antecedentes, los objetivos, los usuarios previstos, las preguntas clave de evaluación, la metodología, las funciones y responsabilidades, los plazos, los entregables, las normas de calidad, las aptitudes del equipo de evaluación y otras cuestiones pertinentes en las que se fundamenta un contrato de UNICEF con los evaluadores.</i>