

ESTUDIOS

La utilidad de las evaluaciones en las ONGD españolas: un estudio basado en la convocatoria de convenios AECID 2006 y 2007

José María Larrú¹ y María Méndez²

Resumen: Mediante una combinación de métodos cuantitativos y cualitativos, el estudio analiza la utilidad de las evaluaciones externas intermedias que las ONGD españolas han realizado de los convenios firmados con la AECID en la convocatorias 2006 y 2007.

La investigación se organizó en torno a cinco grupos de variables exógenas: organización y promoción de la evaluación; participación durante el ciclo evaluativo; el equipo evaluador; utilidad y credibilidad; y difusión de la evaluación.

Los principales hallazgos son que la utilidad y el valor de las evaluaciones percibido por las ONGD es notable (moda de 7 sobre 10) y se explican por la composición del equipo evaluador, la comunicación con dicho equipo y la correspondencia entre lo prometido inicialmente y lo finalmente entregado. Cualitativamente destaca la participación de la ONGD, no la del socio local, ni los recursos invertidos (humanos y de tiempo y dinero). Cualitativamente se identifican varios puntos débiles que conducen a recomendaciones para el futuro.

Palabras Clave: *convenios AECID, evaluación, métodos cuantitativos, análisis cualitativos, ONGD.*

Fecha de recepción: 22 de enero de 2012.

Fecha de admisión definitiva: 12 de julio de 2012.

¹ Departamento de Economía Aplicada. Universidad CEU San Pablo Madrid.

² Consultora y evaluadora independiente.

The usefulness of evaluations in spanish NGO's: a study based on the convocation of AECID agreements 2006 and 2007

Abstract: Combining qualitative and quantitative methods, the study examines the usefulness of the intermediary external evaluations carried out by Spanish NGO's on the agreements signed with the AECID in 2006 and 2007.

The research is organized around five groups of exogenous variables: the organization and promotion of the evaluation, participation in the evaluation cycle, the evaluation team; dissemination of the evaluation.

The main findings are that the usefulness and value of assessments found by the NGO's are remarkable (7 out of 10) and that this is explained by the composition of the evaluating team, this team's communication, and the correspondence between what was initially promised and what was finally delivered. What is highlighted is the qualitative participation of the NGO's, and not that of the local partner or of the investment in resources (whether human, or with respect to time or money). Qualitatively, several weak points are identified, which leads to recommendations for the future.

Key words: AECID agreements, evaluation, quantitative methods, qualitative analysis, NGOs.

L'utilité de l'évaluation des ONGD espagnoles: une étude basée sur les conventions AECID 2006 et 2007

Résumé: À travers une combinaison de méthodes quantitatives et qualitatives, cette étude analyse l'utilité des évaluations extérieures intermédiaires que les ONGD ont réalisé des conventions signées avec la AECID en 2006 et en 2007.

Cette recherche est organisée autour de cinq groupes de variables exogènes : organisation et promotion de l'évaluation ; participation pendant le cycle évaluatif ; l'équipe évaluateur ; diffusion de l'évaluation.

Les principaux résultats sont que l'utilité et la valeur des évaluations perçues par les ONGD sont notables (mode 7 sur 10) et sont expliquées par la composition de l'équipe évaluative, la communication avec cette équipe et la concordance entre ce qui a été promis initialement et ce qui a été finalement rendu. Qualitativement, nous pouvons remarquer la participation de l'ONGD, pas celle de l'associé local ni des ressources investies (personnes, temps et argent). Qualitativement, nous identifions quelques points faibles qui sont à l'origine des quelques recommandations pour l'avenir.

Mots Clés: accords AECID, évaluation, méthodes quantitatives, analyse qualitatif, ONG.

I. Introducción y metodología

El objetivo de esta investigación es conocer si las evaluaciones externas intermedias de los Convenios de la Agencia española de Cooperación Internacional para el Desarrollo (AECID) con las Organizaciones No Gubernamentales de Desarrollo (ONGD) durante las convocatorias de 2006 y 2007 han resultado útiles y las recomendaciones surgidas de las evaluaciones se han incorporado al ciclo de intervenciones (proyectos, programas o nuevos convenios) de las ONGD³. Los Convenios con ONGD son un instrumento de cofinanciación plurianual que la AECID establece bajo concurso público con aquellas ONGD que previamente han superado un proceso de acreditación, bien como socios “generales” o específicos para un determinado sector de intervención (por ejemplo, salud, educación o agua, entre otros). Este instrumento comenzó a funcionar en 2006 y una de sus características diferenciales fue la obligatoriedad de que las ONGD debían realizar una evaluación intermedia externa. Los resultados y efectos de éstas evaluaciones sobre las ONGD no habían sido evaluados hasta la fecha, lo que muestra la relevancia e interés de esta investigación. Ha de advertirse que el objeto de la investigación no son los impactos que generan los convenios sobre las realidades de los países en desarrollo, tampoco sobre la AECID, sino sobre la cultura evaluativa de las ONGD.

La hipótesis principal de esta investigación evaluativa ha sido identificar si las ONGD que debían haber evaluado sus convenios con AECID hasta la fecha (AECID 2009a), habían realizado la obligada evaluación externa e intermedia, cómo la habían llevado a cabo y sobre todo si ésta había sido útil; verificando dicha utilidad en la incorporación de las recomendaciones generadas en el proceso evaluador.

Complementariamente, se pretendía comprender en profundidad la política de evaluación llevada a cabo por las ONGD, para diagnosticar cuáles habían sido los cuellos de botella y fortalezas de dicho proceso, con el fin de poder establecer algunas recomendaciones de mejora de la política de gestión de las evaluaciones de los Convenios AECID-ONGD⁴.

³ Esta investigación fue promovida por el Centro de Estudios de Cooperación para el Desarrollo (CECOD) bajo el título: “La integración de la evaluación en el ciclo del proyecto”, enmarcado dentro del programa cofinanciado por la Comunidad de Madrid: “Fortalecimiento y desarrollo de las capacidades de los agentes de la cooperación madrileña: fortalecimiento institucional y evaluación”, de la convocatoria de 2010.

⁴ La normativa aplicable a esta figura contractual se rige por AECID (2005, 2009a, 2009b).

La evaluación de intervenciones ha sido identificada por la literatura como uno de los medios mejor relacionados con la calidad de la Cooperación para el Desarrollo, tanto por las razones y consecuencias que se derivan de la propia naturaleza no gubernamental de las ONGD (Larrú 2002, 2009) como de las declaraciones internacionales de criterios para una ayuda eficaz (OECD-DAC 2005, 2008, 2011; Larrú 2005).

El estudio se centra en el universo de las 35 ONGD, que debían haber realizado las evaluaciones intermedias obligatorias, en el marco de los Convenios con AECID, de las convocatorias de 2006 (25 ONGD) y 2007 (algunas de 2006 repiten y otras son nuevas). Por lo tanto, no se realizó una metaevaluación de cada una de las evaluaciones, sino que la unidad de análisis e interés en esta investigación es cada ONGD. Se trata de saber cómo cada organización gestiona, sigue, absorbe e implementa los insumos que le aportan el conjunto de evaluaciones que promueve, utilizando como medio de selección la obligatoriedad de evaluar que exige firmar un convenio con la AECID⁵.

Para responder a los objetivos marcados la metodología utilizada fue la siguiente. En primer lugar, se elaboró un cuestionario al que se invitó a responder a las ONGD que se identificaron como universo poblacional con convenios 2006–2007. El cuestionario combinó aspectos cuantitativos y cualitativos. Para obtener información más detallada, se seleccionaron de forma aleatoria 10 ONGD de entre las elegibles que han sido objeto de entrevistas personales en profundidad a partir del cuestionario. Ordenadas alfabéticamente las organizaciones, se generaron números aleatorios en un intervalo (0–1) y las diez organizaciones más cercanas a 0,5 fueron las seleccionadas. Este criterio aleatorio elimina sesgos de selección de entidades por tamaño, orientación ideológica, año de ejecución del convenio u otra característica elegida de forma discrecional por los investigadores.

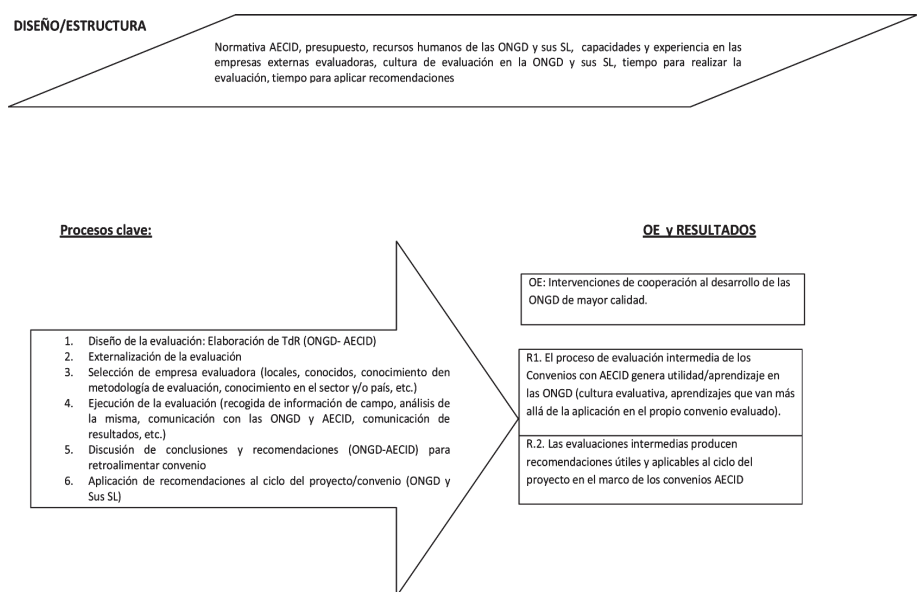
La combinación de los datos cuantitativos (con preguntas cerradas en escalas dicotómicas y ordinales entre 1 y 4, evitando la tendencia central), así como las preguntas abiertas que explican algunas de las respuestas, determinan que las entrevistas sean una herramienta muy válida que evita el positivismo extremo y la pura opinión subjetiva recogida en una entrevista informal. Además, los cuestio-

⁵ En este sentido, esta evaluación no pone su foco de atención en la medición de resultados del sistema evaluado, sino que se centra sobre todo, en la comprensión profunda de la naturaleza del evaluando, en concreto en sus logros y puntos fuertes, así como en sus limitaciones y cuellos de botella. Es decir, se busca la calidad en sentido holístico y se lleva a cabo de manera “comprensiva–respondiente” al tratar de atender los objetivos e intereses de los agentes involucrados (Stake, 2006:85).

narios fueron previamente validados con miembros de ONGD no pertenecientes al universo (sin convenio) lo que permitió depurar y aclarar las preguntas de forma que se limitaran al máximo las preguntas de juicio de valor, de doble sentido o tendenciosas.

En segundo lugar, se realizaron talleres bajo grupos focales entre los miembros de las ONGD participantes lo que permitió una triangulación adicional de las valoraciones obtenidas a partir de los cuestionarios. Por último, se mantuvieron entrevistas en profundidad con el personal responsable de las evaluaciones en la AECID y en la DGPOLDE.

En conclusión, el diseño metodológico ha sido cuidado en extremo y la combinación de métodos cuantitativos y cualitativos permite confiar razonablemente en las valoraciones descritas. La Figura 1 ofrece el análisis sistémico del objeto de evaluación.



El resto del estudio se organiza en otras dos secciones. En la siguiente sección se expone el análisis cuantitativo y cualitativo. La tercera sección resume las principales conclusiones obtenidas.

2. Análisis cuantitativo y cualitativo

En esta sección se presenta el análisis cuantitativo y cualitativo obtenido a partir de la información primaria obtenida por la investigación.

Para el análisis cuantitativo la herramienta de recogida de información fue un cuestionario. El cuestionario diseñado estuvo compuesto por 45 preguntas (o variables). Todas las variables son discretas y cesuradas. El cuestionario se organizó en torno a seis grupos temáticos incluyendo en cada grupo diversa cuantía de preguntas. Los grupos, nombre de las variables y recorrido de cada una se ofrecen en la Tabla 1.

En el análisis previo a la entrega del cuestionario, se identificó un universo de 35 ONGD que habían obtenido convenio en la convocatoria de AECID en los años 2006 y 2007. De este universo se han logrado respuestas de 28 organizaciones lo que representa una muestra del 80%. De las organizaciones de las que no se obtuvo respuesta al cuestionario, tres alegaron que los procesos de las evaluaciones estaban aún en proceso y el resto optaron por no participar.

Para el análisis cualitativo se utilizaron dos herramientas de recogida de información. La entrevista en profundidad realizada a 10 de las ONGD de la muestra, cuyo guión fue la propia encuesta pero formulando las preguntas de la misma de manera abierta. La entrevista se aplicó paralelamente en el tiempo a la encuesta. Posteriormente, y tras un primer análisis de la información obtenida a través del cuestionario y las entrevistas, se desarrolló un grupo de discusión para profundizar en el análisis cualitativo. Para la realización de dicho grupo se invitó a la totalidad de las 28 ONGD que habían contestado al cuestionario, y se obtuvo una participación de 7 ONGD.

A continuación, se comentan los principales resultados obtenidos del análisis de los cuestionarios, entrevistas y grupo de discusión, bajo la perspectiva cuantitativa y cualitativa, agrupados en los seis bloques temáticos.

TABLA I. Grupos, variables y recorrido de las preguntas del cuestionario

	Variable	Recorrido		Variable	Recorrido	
organización y promoción	MOMENTO	1-3	equipo evaluador	Diseño	1-4	
	MOTIVO	1-4		Informac	1-4	
	3OFERTAS	1-4		Obj+Credib	1-4	
	AECID_acepta	1-0		Comunic	1-4	
participación	PO_TdR	1-0		Recomends	1-4	
	PSL_TdR	1-0		completa	1-4	
	PO_preg	1-0		intereses	1-4	
	PSL_preg	1-0		equipo	1-4	
	PO_crit	1-0		ajuste	1-4	
	PSL_crit	1-0		valoración global	Valor	1-10
	PO_campo	1-0		utilidad y credibilidad	Util_proceso	1-4
	PSL_campo	1-0			Creib_concl	1-4
	PO_concl	1-0			Util_recom	1-4
	PSL_concl	1-0			P-mejora	1-0
	PO_recom	1-0	cambiosAECID		1-0	
	PSL_recom	1-0	recuerda		1-0	
	PO_taller	1-0	Concl-útil		1-0	
	PS_taller	1-0	Más-útil		1-5	
	Ev_pregO	1-0	Limit-dific		1-0	
	EV_pregSL	1-0	tiempo		1-4	
PO_informe	1-0	RRHH	1-4			
PSL_informe	1-0	dinero	1-4			
		difusión	difundida	1-0		

Elaboración propia.

2.1. Organización y promoción de las evaluaciones

El cuestionario/entrevista se iniciaba con cuatro preguntas relativas a la organización y promoción del proceso evaluativo. Al ser la primera vez que se emprendían evaluaciones intermedias de los convenios resultaba interesante conocer si *el momento*, que venía determinado por la propia normativa, de la evaluación fue el más oportuno (o hubiera sido mejor otro como al final del convenio o algún tiempo antes/después de la mitad del proceso cuatrienal). También se dio la opción de responder que no se hubiera hecho de no haber sido obligado.

El promedio se situó en 1,85 (recorrido 1–3) con una desviación típica de 0,45. La mayoría de las organizaciones responde que se hizo en el momento oportuno (21 de las 27), cinco opina que hubiera sido más interesante en otro momento y sólo 1 revela que realizó la evaluación por obligación. La explicación cualitativa de esta opción reveló que el convenio comprendía 15 países y en muchos de ellos el proceso planteado es muy largo (incluye segunda fase) y hubiera sido mejor hacerla al final. Esta opción, de poder haber evaluado algo más tarde debido al proceso de desarrollo o al inicio real de las intervenciones más allá de la fecha formal, es la razón más frecuente tras la selección de “hubiera sido más interesante realizarla en otro momento”. A modo de ejemplo, se cita una frase textual recogida de la información cualitativa:

la realidad es que estaban pensados como convenios de mayor duración (todos ellos tienen segunda fase) por lo que hubiera sido más útil una evaluación final que en realidad hubiera sido intermedia del proceso tal y como estaba pensado.

De la información cualitativa, también se desprende que el proceso de evaluación completo es bastante largo, mínimo cinco o seis meses desde que se comienza a elaborar los Términos de Referencia (TdR de aquí en adelante) hasta que se da por concluido el informe, lo que hace que apenas quede un año de ejecución para aplicar las recomendaciones, muchas de las cuales son demasiado ambiciosas para aplicarlas en lo que resta del convenio.

El *motivo* predominante al realizar la evaluación ofrece mayor dispersión (desviación típica de 0,97 para un promedio de 2,7). Dado que se abría la posibilidad de respuesta múltiple hay que interpretar con cautela el promedio. La respuesta con frecuencia más alta es que la evaluación se realiza para establecer un proceso de *mejora continua* (se señala en 16 ocasiones, el 42% de las observaciones), seguida por analizar los *procesos internos* (9 veces, 24%) y conocer los *resultados* logrados (8 veces). Cumplir con la *obligación* marcada por el convenio se señaló 5 veces (el 13%). Resulta interesante advertir que, de la información facilitada por

las entidades, no se desprende una dicotomía entre la evaluación de procesos y la de resultados, obteniendo ambas casi la misma selección.

De la información cualitativa se desprende que el modelo de TdR que es obligatorio utilizar actúa como un corsé. Los siguientes testimonios lo muestran con claridad:

Más allá de los cinco criterios del CAD nos interesaba evaluar los procesos, pero conociendo a la AECID no te vas a meter a añadir cosas. Nosotros directamente les dijimos a los evaluadores donde nos interesaba que se centraran más allá de lo que reflejaban los TdR.

Quisimos meter un criterio más y fue todo un lío.

Otra debilidad era que el objetivo de la evaluación era demasiado amplio, los 5 criterios, mas procesos, etc. Hubiera sido más interesante haberse centrado en alguna cuestión concreta porque por ejemplo, los procesos no fueron bien analizados y hubiera sido más interesante en este momento.

El informe no se ajustó a los 5 criterios y AECID nos hizo reordenar todo el informe de nuevo. Para nosotros fue tiempo perdido pues el informe nos parecía bueno ya que respondía a lo que a nosotros nos interesaba.

Varias organizaciones afirman que trataron de no meter algún criterio de los 5 recomendados, pero la AECID les obligó a incorporarlo⁶.

Incluso más allá de los propios TdR, las ONGD opinan que depende del técnico responsable del Convenio por parte de AECID y de la OTC correspondiente el tener mayor o menor flexibilidad para centrar la evaluación en lo que realmente le interesa a la ONGD.

Esta participación de la AECl para reflejar sus intereses en el proceso de evaluación, ha implicado una gran participación en la fase de elaboración de los TdR. En este sentido, se recogen algunas citas de la opinión que las ONGD tienen sobre este proceso:

El tiempo empleado por nuestro personal ha sido excesivo sobre todo en la negociación con AECID de los TdR que se meten a opinar hasta el último detalle de la logística (por ejemplo

⁶ Los cinco criterios estándar de evaluación del modelo de la OCDE son: pertinencia o medida en que los objetivos son congruentes con los requisitos de los beneficiarios; eficacia (grado en que se lograron los objetivos propuestos); eficiencia (medida en que los recursos/insumos se han convertido económicamente en resultados); impactos (efectos de largo plazo, positivos y negativos, primarios y secundarios, producidos directa o indirectamente por una intervención, intencionalmente o no); sostenibilidad (continuación de los beneficios de una intervención después de concluida). Véase el glosario de la OCDE (2002).

les parecía mucho tiempo de trabajo de campo y poco de revisión de documentación), este tipo de observaciones, no aportan mucho, casi lo llamaríamos de "intromisión", y hace perder muchísimo tiempo.

En todos los casos, el tiempo empleado en la elaboración de los TdR fue excesivo debido al nivel de detalle en que AECID hacía sus aportaciones y exigencias, la discusión de los TdR fue un proceso interminable y agotador.

Los TdR se manipulan mucho, de aquí para allá y se acaba perdiendo el enfoque de lo esencial, de lo importante.

La pregunta sobre cómo fueron seleccionadas las empresas evaluadoras también permitió una selección múltiple. Se han obtenido 44 observaciones. La más frecuente es el contacto con empresas de las que ya se tiene buenas referencias (15 selecciones que representan el 34%). Prácticamente al mismo nivel se sitúa el concurso público (14 observaciones, el 32%). La posibilidad de recibir ofertas de empresas con las que ya se había trabajado anteriormente obtuvo 8 observaciones y 7 la opción de recibir ofertas de las que no se tenían referencias previas. Las elecciones combinadas se reparten del siguiente modo: 3 entidades optan por la combinación 2-3 (empresas con contacto previo y de las que se tenían buenas referencias por otras organizaciones); dos veces la combinación 2-3-4 que añade a la anterior empresas sin referencias previas; una vez la 1-3-4 y una vez 1-2-4 y una entidad que marcó todas las respuestas posibles. Como conclusión, puede señalarse que las ONGD, dado que les es obligado hacer la evaluación externa, optan por hacer un concurso pero con cierta "restricción" a empresas con las que ya se ha trabajado o de las que se han obtenido buenas referencias.

Desde un punto de vista de la información cualitativa recogida, al responder a esta pregunta, inevitablemente surgió la opinión de las ONGD sobre la pertinencia de la "obligatoriedad de evaluaciones", lo cual se refleja a través de las siguientes citas:

De igual modo, pienso que se es muy estricto con el carácter externo de las evaluaciones, ya que en algunos casos si las evaluaciones son mixtas⁷ pueden conseguirse mejores resultados para todos.

Un equipo mixto agregaría valor añadido.

Si lo que se está buscando es aprendizaje entonces sería mejor una evaluación mixta, pero si lo que se busca es una rendición de cuentas entonces creo que es mejor que sea externa e incluso debería ser contratada por la propia AECID para garantizar mayor transparencia.

⁷ Una evaluación mixta es la que incorpora agentes de la propia unidad que es evaluada al equipo externo.

En una evaluación intermedia lo principal es aprender para corregir y por ello sería mucho mejor una evaluación mixta.

Una mención aparte ha sido la introducción en la evaluación de los convenios de 2007 de que una de las ofertas tenía que ser de una empresa local.

Algunos testimonios sobre la opinión de las ONGD con respecto a este punto son:

Los evaluadores locales tenían menos experiencia y no llegaban a comprender la metodología impuesta por AECID, los criterios, etc. Se les quedó grande el trabajo, además de no respetar los tiempos marcados, no se respetaron los TdR.

En todos los casos nos obligaron a presentar alguna oferta de empresa evaluadora local. Sobre este tema no le encontramos mucho el significado porque en convenios multi-país, el hecho de que el evaluador sea boliviano, por ejemplo, a veces es peor que si fuera español, para evaluar en Perú, por la rencilla cultural que hay entre estos dos países, y por otro lado, ¿por qué el boliviano va a conocer mejor la cultura de Perú?

Se contrató una empresa española que tenía personal local en cada país del convenio pero al final el resultado fue que cada evaluador evaluó su país y no se dio una visión regional.

Con respecto a la pregunta sobre si la AECID aceptó la propuesta realizada en la elección de la empresa, todas las ONGD responden que sí excepto una. El motivo fue que al ser de 2007, se solicitó por parte de AECID la repetición del concurso público a nivel local (el convenio se ejecutaba en Colombia). La experiencia no fue buena ya que no se presentó nadie y se volvió a la propuesta original, retrasando mucho el proceso.

2.2. La participación a lo largo del proceso evaluativo

Esta sección del cuestionario se compuso de 9 ítems que preguntaban de forma dicotómica (por tanto se da el valor 1 si se contestó con un "sí" y valor 0 si la selección fue un "no") sobre nueve momentos en los que se dividió el ciclo de la evaluación. Desde los términos de referencia a la participación en los momentos anteriores a la recepción del informe de evaluación. Para aligerar el análisis, hemos procedido a agregar las respuestas obtenidas en las que sí se participa por parte de la ONGD y –en otro grupo– la participación de los socios locales. Dos entidades sólo respondieron a la participación de la ONGD, con lo que éstas tienen más respuestas que la de los socios locales. Los resultados de forma resumida se ofrecen en la tabla 2.

TABLA 2. Participación a lo largo del proceso evaluativo: frecuencias acumuladas

	TdR	Preguntas	Criterios	Trabajo de campo	Conclusiones	Recomendaciones	Taller en terreno	¿Pidió el evaluador?	Antes de recibir el informe
ONGD	27	27	12	16	18	15	22	26	27
Socio Local	14	15	*2	15	14	8	18	20	19

Nota: el número de respuestas de ONGD es 28, mientras que el de Socio Local son 26, excepto en el ítem "criterios" que es 25.

Elaboración propia.

Como se desprende de la tabla, prácticamente todas las ONGD han participado en la elaboración de los términos de referencia, en la formulación de las preguntas evaluativas y en la fase previa a la recepción del informe final de evaluación. La frecuencia más baja de participación es en la selección de los criterios de evaluación (12 de 28). No contar con la libertad por parte de AECID para cambiar los cinco criterios canónicos del CAD está tras esta menor participación. También la participación desciende en el momento del trabajo de campo, elaboración de las conclusiones y de las recomendaciones. La interpretación que se hace de esta menor frecuencia es que se respeta la independencia de los evaluadores al ser una evaluación externa.

Por lo que se refiere a la participación de los socios locales en las evaluaciones, vemos que es menor que la de las ONGD. Mientras las ONGD ofrecen una participación global de 190 respuestas positivas ("sí") el valor del socio local es de 123 síes. La *menor participación de los socios* se produce en la selección de criterios (2), seguida de la elaboración de las recomendaciones (8). La participación es de casi el doble cuando se hace referencia a las fases finales del ciclo evaluativo: taller en terreno, solicitudes del evaluador o antes de recibir el informe final (fase de borrador del informe).

Es llamativo que en la mitad de los casos posibles, los socios no participan en la elaboración de los términos de referencia ni en la formulación de las preguntas evaluativas. De hecho, las mayores desviaciones típicas dentro del grupo de variables relativas a la participación se producen con los socios locales (términos de referencia, preguntas, campo y conclusiones superan el 0.50 de desviación típica).

La información recogida de manera cualitativa muestra que las ONGD consideran fundamental participar en el proceso de evaluación:

Lo fundamental es participar desde el principio porque si no le dedicas el tiempo suficiente desde el primer momento, luego te encuentras con los problemas.

Cuando participas todo fluye mucho mejor y tienes la capacidad de negociar aunque estés a mitad del proceso.

Si hay participación y un buen equipo de evaluación, todo lo demás viene dado.

De hecho, las 7 ONGD participantes en el grupo de discusión afirmaron que la participación es uno de los elementos clave para el desarrollo de buenas evaluaciones.

Del análisis de la información recogida de manera cualitativa se desprende que para las ONGD con mayor cultura evaluativa (las que ya cuentan con personal específico formado en la materia o cuentan con un plan de evaluaciones, por ejemplo), el aspecto de la participación es fundamental e incluso consideran que el aspecto clave en todo el proceso es saber formular bien las preguntas de evaluación, transmitir lo que se quiere conocer de manera adecuada al equipo de evaluación.

Para las ONGD con menor cultura evaluativa, incluso para aquellas en las que estas evaluaciones eran las primeras que se realizaban en la organización, también se reconoció el aspecto de la participación como clave:

La evaluación ha sido muy útil, en primer lugar por la posibilidad de acompañar el proceso de evaluación, siendo el aprendizaje muy interesante para la organización.

Se han sacado muchos aprendizajes sobre todo de cómo hacer un proceso de evaluación, ha sido compartido con toda la organización.

Incluso, como veremos más adelante cuando analicemos los principales cuellos de botella, se verá como algunos de ellos se podrían haber superado con una mayor participación por parte de las ONGD.

Sin embargo, con respecto a la participación de los Socios Locales (SL a partir de ahora), corroborando los datos del análisis cuantitativo, se aprecia una menor participación en las fases iniciales del proceso: TdR, definición de preguntas de evaluación y de los criterios. Las ONGD, a través de la información cualitativa recogida, expresan que los socios locales tienen menos cultura de evaluación y que ésta es considerada como un mero trámite a cumplir, sin que exista un mayor interés por el aprendizaje o la mejora. Según las ONGD éste es el motivo principal que explicaría la menor participación de los SL en todo el proceso de evaluación:

Los socios locales lo siguen viendo como una imposición, nosotros ya lo vemos como una oportunidad.

2.3. Los equipos de evaluadores externos y su trabajo

El tercer grupo de preguntas tuvieron la finalidad de conocer la percepción que las ONGD tenían sobre el desempeño de los equipos de evaluación (o evaluador individual). Se diseñaron nueve ítems y se ofrecía una escala ordinal de dimensión cuatro (para evitar la tendencia central). Las opciones de respuesta eran malo–regular–bueno–muy bueno. La codificación se ha realizado en números naturales positivos, dando valor 1 a malo y 4 a muy bueno.

La organización del análisis de este grupo se hace en dos etapas. Primero procedemos a una visión de conjunto de las 9 variables y posteriormente analizamos una por una.

Todas, excepto la de diseño, tienen 28 observaciones (respuesta completa). El diseño, la recogida de la información, la valoración del equipo y el ajuste entre trabajo diseñado y realmente efectuado tienen un valor mínimo de 1 (malo). El mínimo del resto de variables es 2. El máximo es 4 en todas las variables.

La frecuencia del valor 1 (“malo”) es 4 (el 2% de las 239 observaciones que componen el total del grupo sobre el trabajo de los evaluadores). La de 2 (“regular”) es 62 (26% de las respuestas). El valor 3 (“bueno”) se eligió en 151 ocasiones, representando el 63%. La opción 4 (“muy bueno”) obtuvo 22 observaciones (el 9%). En resumen, la opinión predominante en las ONGD con convenio ha sido que el trabajo realizado por los evaluadores ha sido “bueno”, aunque la cuarta parte se decanta por regular (26%).

La mayor satisfacción se concentra en la *comunicación* mantenida entre evaluadores y ONGD (7 frecuencias de “muy bueno” que son el 26%, muy por encima del 9% que tiene ese valor en el global). Por encima del valor promedio de la calificación “bueno”, están las variables “*diseño*” (frecuencia 20 que es un 74%); “*ajuste*” (frecuencia 19 que es un 73%); “*equipo*” (con frecuencia 18 que es un 69%) y “*recomendaciones*” (con 18 observaciones sobre 26 –una menos que “equipo”, lo que supone un 67%).

Por el lado de la valoración “regular”, sobresalen “*completa*” (11 observaciones que es un 42%) y la *información recogida* (10 que supone un 38%).

En resumen, las ONGD expresan satisfacción con la comunicación mantenida por los equipos evaluadores y, en menor medida, con el diseño, ajuste entre el trabajo finalmente realizado y el diseñado, el equipo y las recomendaciones recibidas. En la valoración cualitativa se detecta que hay varias quejas sobre las recomendaciones en el sentido de que son muy generales, ya sabidas o muy difíciles de implementar en la práctica:

Las recomendaciones emitidas por los evaluadores son, con frecuencia, poco aplicables. Al basarse en una observación puntual y un conocimiento del contexto específico con limitaciones, en ocasiones las recomendaciones resultan un tanto teóricas y poco adaptadas a la realidad. Y en la mayor parte de los casos no aportan el valor añadido de asistencia técnica que se le supone a una evaluación, ya que no se aportan los "cómos" a los "qués". En general las evaluaciones son muy útiles para poner de relieve "qué" se debe hacer o corregir, pero raramente aportan "cómo" hacerlo o corregirlo en el contexto real concreto, por lo que la puesta en marcha de las propuestas de mejora es difícil.

Las propuestas de mejora se quedaron escasas y el documento era demasiado extenso (372 páginas), lo que hace que sea poco práctico.

Donde menos satisfacción parece haber (o mayor diferencia de opiniones) es la cantidad de información recogida (relacionada directamente con el tiempo pasado en el terreno), tal y como se recogen en los siguientes comentarios:

La principal limitación es no tener tiempo suficiente para el levantamiento de la información y eso de lugar a evaluaciones de poco rigor.

Es poco el tiempo que se pasa en terreno para evaluar un proyecto y entender la ejecución del mismo, conocer actores importantes, etc.....

También hay menor satisfacción en que la evaluación haya resultado completa, identificando fortalezas y debilidades. La información recogida de manera cualitativa lo corrobora a través de comentarios como:

Se ensañaron en fallos que incluso no representaban ni el 1% del presupuesto.

Hay una tendencia a destacar sobre todo las debilidades.

El informe fue duro, solo se centró en las debilidades.

Pasando ahora al análisis variable por variable, los estadísticos descriptivos revelan que, en promedio, la mayor satisfacción se obtiene en la *comunicación* (media de 3,05 y desviación típica de 0,67 la mayor de entre los nueve ítems). A continuación, el *diseño* obtiene un promedio de 2,85 puntos (y dispersión de 0,60). Le siguen *objetividad/credibilidad* de la interpretación de la información recogida, las *recomendaciones*, y que se hayan tenido en cuenta los *intereses* de la ONGD. Las tres tienen un promedio de 2,80 y desviación típica de 0,61. Un cuarto grupo

lo constituyen la valoración promedio del equipo y el ajuste a lo propuesto, con una media de 2,78 y desviación típica de 0,61 y 0,64 respectivamente. Por último, obtienen un promedio de 2,57 la información recogida (dispersión de 0,62) y la completitud de la evaluación (dispersión de 0,48).

Para una visualización más amable de este "ranking" la tabla 3 contiene el promedio y desviación típica de los ítems ordenados de mayor a menor media.

TABLA 3. Ranking del promedio de los ítems relacionados con la calidad del equipo evaluador

Variable	Promedio	Desviación típica
Comunicación	3,0536	0,6713
Diseño	2,8519	0,6015
Objetividad-Credibilidad	2,8036	0,6137
Recomendaciones	2,8036	0,5500
Intereses	2,8036	0,6137
Equipo	2,7857	0,6150
Ajuste	2,7857	0,6444
Información	2,5714	0,6193
Completa	2,5714	0,4852

Elaboración propia.

2.4. La valoración cuantitativa global

La variable con mayor recorrido dentro del cuestionario fue la de ofrecer una valoración global del proceso de evaluación entre 0 (peor) y 10 (mejor valoración). El mínimo ha resultado ser un 3 y el máximo un 9, con un promedio de 6.58 y desviación típica de 1,36.

Como se puede observar en la tabla de frecuencias (Tabla 4), la valoración más elegida ha sido el 7 (frecuencia 12) agrupando casi el 43% de las respuestas.

TABLA 4. Frecuencias de la valoración global

		Valor			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	3,00	1	3,6	3,6	3,6
	4,00	2	7,1	7,1	10,7
	5,00	2	7,1	7,1	17,9
	6,00	4	14,3	14,3	32,1
	6,50	1	3,6	3,6	35,7
	7,00	12	42,9	42,9	78,6
	8,00	5	17,9	17,9	96,4
	9,00	1	3,6	3,6	100,0
	Total	28	100,0	100,0	

Elaboración propia.

Puede concluirse que la satisfacción promedio con el proceso evaluativo es de "notable", lo que puede considerarse alta si se tiene en cuenta que es una experiencia novedosa que se realiza por primera vez.

De la información cualitativa, se recoge alguna información que complementa la anterior y que surgió de manera espontánea al no ser preguntas directamente formuladas en el cuestionario (por ejemplo, cuando se preguntó sobre los principales cuellos de botella en el proceso). Como se podrá observar a continuación todos los comentarios se centraron en el propio equipo de evaluación, por lo que parece ser el tema que más preocupa a las ONGD, y que como veremos más adelante en el análisis cuantitativo, es la variable que más influye en el grado de utilidad de las evaluaciones realizadas. A continuación se detallan los comentarios más relevantes.

En general, las ONGD opinan que es mejor contar con equipos evaluadores compuestos de un mínimo de 2 personas (el máximo de personas en un equipo fue de 5) que con un solo evaluador.

También existe una queja generalizada cuando en el proceso se contrata a una empresa que libremente cambia el equipo evaluador que inicialmente se había comprometido. Las ONGD consideran que la clave de la calidad del equipo

evaluador está en las personas y no en la empresa porque incluso en la misma empresa hay personas que trabajan mucho mejor que otras:

En algún caso la persona que ha realizado la evaluación no tenía buena preparación, en ese caso se contrató con empresa consultora unos TdR y luego ellos buscaron a la persona evaluadora y la propia empresa consultora reconoció el desastre de la evaluación.

La explicación de los malos resultados se debe a que en los TdR la empresa ganadora ofrecía un determinado personal que luego no se respetó.

Al final, los resultados dependen mucho de la aptitud y actitud de los evaluadores.

También surge repetidamente un dilema entre la contratación de expertos en evaluación y/o la contratación de expertos en el sector y/o país en el que se desarrolla la intervención. En este sentido:

Se dio un choque de teorías entre el evaluador, que además era experto en su tema (educación, o agrario) y el propio responsable del SL que era experto en esa misma materia. Al final, el informe parece más un informe de experto en la materia (con recomendaciones muy útiles) pero no tanto un informe de evaluación, pierden un poco el rol de evaluador para decantarse por el rol de experto en la materia.

Es muy difícil encontrar gente que sea buena evaluadora y a la vez expertos en el tema, por eso la importancia de los equipos mixtos porque la ONGD puede aportar la experticia en la materia.

Según los diversos comentarios y tal y como se expresó en el grupo de discusión, parece que la mejor opción es decantarse por expertos en evaluación y complementar sus limitaciones en conocimientos concretos en el tema a través de la propia ONGD o mediante la creación de equipos.

Por último, otro tema que salió recurrentemente en la información cualitativa sobre el equipo evaluador es la dificultad de encontrar verdaderos expertos en evaluación en muchos de los equipos que se contrataron. Esta fue apuntada como la principal debilidad de los equipos de evaluación. Las ONGD echan en falta experiencia tanto en el diseño de la propia metodología de evaluación como en las técnicas de recogida de información. Algunos comentarios lo muestran:

Muchas veces los datos no son fiables porque van allí y apuntan lo que les dice el primero que se encuentran y no hacen una adecuada interpretación.

Es difícil saber cuál es la mejor propuesta técnica porque todas las metodologías eran super-pobres.

Y hablar de técnicas cualitativas y cuantitativas, eso ya es ciencia ficción, ja, ja...

2.5. Utilidad y credibilidad

Este grupo se estructuró en torno a cuatro bloques de conocimiento. Primero se interrogó sobre la utilidad general del proceso, luego sobre las consecuencias concretas que ha tenido la evaluación, en un tercer momento para quién ha resultado más útil y en cuarto lugar cómo se percibía el tiempo, recursos humanos y dinero empleados en la evaluación.

Sobre la *utilidad* general de la evaluación se obtienen los siguientes resultados. Ninguna ONGD ha valorado el proceso con la puntuación mínima ("nada útil"), predominando con la misma frecuencia (13 observaciones) la valoración "algo" y "mucho" o muy útil (las dos máximas puntuaciones que había en la escala de 4). Ha habido un caso de "poco" y una indecisión entre "algo-mucho" (codificada con un puntaje de 3,5). El promedio es 3,44 con desviación típica de 1,36. Es decir, el 92% de las entidades valora el proceso como útil o muy útil. Este dato cuantitativo se corrobora con la información recogida de manera cualitativa, tal y como lo muestran comentarios del tipo:

La evaluación ha sido muy útil, en primer lugar por la posibilidad de acompañar el proceso de evaluación, siendo el aprendizaje muy interesante.

Nos ha aportado mucho para mejorar los convenios que estamos ejecutando en la actualidad.

Sobre la credibilidad de *las conclusiones y juicios*, los resultados son muy parecidos. Con un promedio de 3,41 ($\sigma=0,62$) hay 13 valoraciones de "mucho" y 12 de "algo" (es decir, 25 valoraciones máximas en la escala de 4).

Un poco menos útiles han sido valoradas las recomendaciones, pero con un promedio de 3,34 ($\sigma=0,61$), la opción "algo" es elegida en 14 ocasiones y "mucho" en 11, es decir, 25 ONGD se muestran satisfechas con las recomendaciones obtenidas a pesar de que en la información cualitativa se recogieron algunos comentarios críticos hacia las recomendaciones en el sentido de que eran muy generales, ya sabidas o muy difíciles de implementar e incluso hubo alguna ONGD que mencionó la poca utilidad de las recomendaciones como principal cuello de botella para integrar la evaluación en el ciclo del proyecto.

Es decir, la percepción revelada sobre la utilidad del proceso, las conclusiones y recomendaciones parece ser alta y muy alta en la mayoría de las ONGD.

Esta alta valoración conduce a querer conocer *las consecuencias* concretas que ha tenido la evaluación sobre las intervenciones de las ONGD.

Una primera consecuencia inmediata podía ser elaborar un *plan de mejora* por escrito. Aunque esta acción es obligada según el procedimiento establecido por AECID, 5 organizaciones (el 18%) responden de forma negativa. En las ONGD que fueron entrevistadas se aludió que el motivo era precisamente el momento en el que se encontraba actualmente el proceso de evaluación, es decir, que se acababa de recibir el informe y todavía se estaba analizando o que se estaba en pleno proceso de elaboración. Para el resto de organizaciones, las encuestadas, no se aluden los motivos para no contar con este plan de mejora. Sin embargo, resulta interesante mencionar que en la información cualitativa recogida son muchas las ONGD que afirman que su plan de mejora va más allá de la propia agenda firmada con AECID, tal y como lo muestra comentarios del tipo:

Nuestro plan de mejora incluye una hoja de ruta mucho más exigente que lo que nos pedía la AECID, fuimos recomendación por recomendación aplicando mejoras.

El convenio estaba formado por 6 SL y se han elaborado un plan de mejora para ser ejecutado por cada uno de ellos. El plan de mejora va mucho más allá de lo exigido por la AECID.

Otra consecuencia de la evaluación, al ser intermedia, es que se introduzcan *modificaciones* y se soliciten cambios al financiador (AECID). La distribución de respuestas es que el 46% (13 organizaciones) sí solicitaron cambios y el 54% no (15 entidades). En este punto es importante aclarar que por exigencias de la propia normativa, sí se pueden estar dando modificaciones y plan de mejora a un convenio evaluado pero que no requiera solicitar formalmente un cambio porque no supone modificación sustancial (para conocer en mayor detalle las características de las modificaciones sustanciales puede remitirse a la normativa AECID que regula a los convenios).

Para verificar de forma más segura las consecuencias de la evaluación, se preguntó *si se recordaba alguna recomendación* surgida del informe de evaluación y se dijera cuál. El 89% responde de forma afirmativa, pero 3 ONGD (10,7%) confiesan no recordar en ese momento ninguna recomendación.

Por último, dentro del bloque de las consecuencias, se preguntó si se habían extraído conclusiones útiles para ser aplicados en programas similares de la ONGD (*replicabilidad*). Eligieron la respuesta afirmativa 26 entidades (el 93%).

De la información cualitativa se recogen algunas de estas utilidades, a modo de ejemplo, podemos citar:

Una clara recomendación ha sido centrarse en un tipo de derechos a trabajar con las mujeres y no de manera tan integral como hicimos pues esto provocó una dispersión en el efecto.

Otra organización comentó que se ha realizado una sistematización interna de todas las recomendaciones de las evaluaciones y éstas han sido difundidas dentro de la organización. También se recogen comentarios sobre recomendaciones útiles para el diseño de nuevas intervenciones en relación a su estructura de gestión (integrar menos proyectos, trabajar en menos países en un mismo convenio, reformulación de indicadores para que sean más SMART, mejorar las fuentes de verificación, acotar los sectores de intervención), incluso se han obtenido recomendaciones a aplicar sobre el propio proceso de evaluación para ser aplicadas en futuras evaluaciones (cerrar muy bien desde los TdR los CV de las personas evaluadoras, concretar mucho más las preguntas de evaluación y definir en los TdR de forma muy precisa el producto final que se quiere, el tipo de informe).

Tras la valoración de la utilidad y las consecuencias, se preguntó a qué *agente* se pensaba que había beneficiado más la evaluación y se ofrecían, AECID, la ONGD, el socio local, los beneficiarios u otros. La distribución de frecuencias (se permitió la respuesta múltiple) ha sido la siguiente:

TABLA 5. Frecuencias de la pregunta “¿Para quién cree que ha sido más útil?”

Opción	Frecuencia	%
AECID	7	14%
ONGD	23	47%
Socio Local	13	27%
Beneficiarios	6	12%
Suma	49	100%

Elaboración propia.

El ranking de elecciones es ilustrativo. La evaluación parece haber sido más útil a la ONGD (sujeto que responde) seguido de las contrapartes locales. AECID y beneficiarios son escogidos casi en la misma cuantía. Las razones cualitativas de esta elección son amplias. Las ONGD pueden considerar que los beneficiarios obtienen fruto de la evaluación de manera indirecta, en la medida en que se mejoran las intervenciones, pero no de manera directa porque consideran que todavía siguen siendo sujetos pasivos a los que se les hace un “feedback” del proceso como mucho.

Con respecto a la AECID, las ONGD consideran que ésta no tiene capacidad para asimilar todo el volumen de evaluaciones realizado.

Consideran también que las principales recomendaciones van en la línea de comprender mejor la propia intervención, mejorar los procesos y mejorar los sistemas de recogida de información y en ese sentido son las ONGD y los SL los responsables de esta tarea.

Respecto a las limitaciones señaladas, tras haber respondido de forma afirmativa a la pregunta de si hubo *condicionantes* importantes para alcanzar los resultados previstos en la evaluación se analizaron las posibles causas. De las 28 ONGD que responden, 20 perciben que sí hubo limitaciones (el 71,4%) mientras que 8 responden negativamente. La mayoría de los condicionantes se ha ido recogiendo a lo largo del texto cuando se han analizado los diversos apartados, sin embargo, a modo de ejemplo, mencionamos aquí algunos: la falta de dominio del idioma local (quechua), el momento era el marcado por la normativa pero no era el adecuado, la metodología no es la adecuada para un convenio regional (se dan recomendaciones por país pero no por región), el hecho de que se obligue a evaluar bajo los criterios del CAD no permite dar tanta importancia a los procesos que en una evaluación intermedia debería ser el elemento central, el tiempo que se pasa en terreno es muy escaso, entre otros.

El último bloque en torno a la utilidad-credibilidad de la evaluación gira en torno a la percepción de los recursos empleados. Bajo una escala ordinal de cuatro elementos (insuficiente-suficiente-proporcional-excesivo) codificada entre 0 (insuficiente) y 4 (excesivo) se preguntó sobre el tiempo empleado, los recursos humanos (RR.HH.) involucrados y el dinero invertido.

Los estadísticos de tendencia central ofrecen valoraciones muy similares: el promedio de tiempo es 2,46 (con $\sigma=0,92$); de RR.HH. es 2,60 ($\sigma=0,73$); y de dinero 2,55 ($\sigma=0,97$). Las medias de dispersión son, quizá, más informativas que el promedio. Hay más diversidad de opinión en valorar el dinero, seguida del tiempo. Más concentrado aparece el juicio sobre los RR.HH.

Respecto al *tiempo*, la distribución de frecuencias contiene 4 casos en los que se optó por los valores más extremos. Cuatro ONGD consideran que el tiempo fue excesivo y cuatro que fue insuficiente. La mayoría (39,3%) lo valora "suficiente", seguido muy de cerca por "proporcional" (32,1%). La información cualitativa recogió algunas quejas sobre la limitación del tiempo que habitualmente los evaluadores emplean en el trabajo de campo, lo que limita mucho la aplicación de diversas

técnicas para la recogida de información y una organización también subrayó la falta de tiempo dedicado al trabajo de gabinete.

Con respecto al tiempo también existe la percepción generalizada de que es escaso no para realizar la evaluación pero sí para procesar el aprendizaje obtenido y para implementar mejoras al quedar poco tiempo de vida del convenio una vez obtenido el informe de evaluación.

Por último, también se manifiesta de forma generalizada que el tiempo empleado por el personal ha sido excesivo en la negociación de los TdR con AECID.

TABLA 6. Tiempo

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 (insuficiente)	4	14,3	14,3	14,3
	2,00 (suficiente)	11	39,3	39,3	53,6
	3,00 (proporcional)	9	32,1	32,1	85,7
	4,00 (excesivo)	4	14,3	14,3	100,0
	Total	28	100,0	100,0	

Elaboración propia.

Con respecto a los recursos humanos, se mantiene la proporción de las organizaciones que valoran el esfuerzo como "excesivo", pero hay una dominancia notable en torno a "suficiente" (el 53,6%).

TABLA 7. Recursos Humanos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,00 (suficiente)	15	53,6	53,6	53,6
	3,00 (proporcional)	9	32,1	32,1	85,7
	4,00 (Excesivo)	4	14,3	14,3	100,0
	Total	28	100,0	100,0	

Elaboración propia.

Por último, respecto a los *recursos monetarios*, se mantiene el número de valoraciones de insuficiente en cuatro casos, excesivo acumula casi el 18% (5 casos), y la dominancia se centra entre suficiente 28,6% y proporcional (35,7%). No parece que la percepción del coste de las evaluaciones haya sido excesiva, aunque de forma cualitativa sí hay testimonios de que el presupuesto para llevar a cabo las evaluaciones fue reducido. De forma cualitativa hubo quejas de escasez de recursos económicos, pero relacionadas con la cantidad de preguntas evaluativas formuladas, fruto de no poder seleccionar (reducir) los criterios de evaluación (los cinco del CAD), de hecho algunas ONGD optaron por realizar evaluaciones finales con recursos propios y éstas fueron dotadas de un mayor presupuesto que las intermedias.

TABLA 8. Dinero invertido

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 (insuficiente)	4	14,3	14,3	14,3
	1,50*	1	3,6	3,6	17,9
	2,00 (suficiente)	8	28,6	28,6	46,4
	3,00 (proporcional)	10	35,7	35,7	82,1
	4,00 (excesivo)	5	17,9	17,9	100,0
	Total	28	100,0	100,0	

Nota: la valoración 1,5 puntos fue asignada al cuestionario que optó por responder insuficiente-suficiente sin discriminar hacia una de ellas.

Elaboración propia.

En resumen, no se deduce de las respuestas una falta de recursos que se correlacione con la utilidad-credibilidad de la evaluación, sin embargo, recogemos un testimonio que agrupa la percepción de los tres tipos de recursos utilizados:

El dinero para hacer la evaluación, determina la composición del equipo evaluador, y el tiempo que se emplea para recoger información en terreno, son dos cuestiones básicas para hacer un trabajo de calidad.

2.6. Difusión

La última sección del cuestionario preguntó sobre la difusión del informe de evaluación. El medio que se proponía era la página web de la ONGD y se daba

libertad para revelar otros. De las 28 ONGD, sólo 6 (21%) han respondido que han difundido de alguna forma su evaluación. Este tema fue ampliamente debatido en el grupo de discusión y se obtuvieron las siguientes explicaciones al porcentaje tan bajo de difusión recogido en el análisis cuantitativo.

Parece que hay un común acuerdo que cultura evaluativa organizacional y difusión van de la mano. En opinión de la mayoría de las ONGD, esta cultura evaluativa se está generando (y en muchos casos a raíz del mecanismo de evaluación inducido por AECID a través de la normativa de obligatoriedad para los convenios) y consecuencia de ello es que hay varias ONGD (precisamente las que parecen tener mayor cultura evaluativa: personas contratadas exclusivamente para labores de evaluación, personal formado específicamente en el tema, mayor número de evaluaciones realizadas, etc.) que ya están integrando la difusión de las evaluaciones en sus planes organizativos y como tema de reflexión institucional. Algunos comentarios reflejan este hecho:

Antes ni nos lo planteábamos, ahora con la cultura evaluativa que se está implantando ya sí se ha recogido en el Plan Estratégico.

Era algo nuestro, hoy en día nos hemos propuesto tener un plan de difusión para 2012).

Si los convenios se ejecutan con fondos públicos y las evaluaciones también, éstas deberían ser un bien público.

El departamento de evaluación es nuevo, de reciente creación, y todavía estamos elaborando los procedimientos y precisamente uno de ellos es la difusión y el uso de las evaluaciones.

En línea con este último comentario, también hay ONGD que muestran que la falta de cultura evaluativa en sus SL les limita para poder publicar las evaluaciones ya que estos últimos no autorizan a ello:

Se quiso publicar una evaluación en la red de los propios SL porque había sido muy buena, pero el propio SL se opuso a ello.

Sin embargo, también se recogieron comentarios que reflejan el temor a publicar las evaluaciones.

Es un problema estructural del propio sistema. En un sistema que dependes de la subvención no te puedes arriesgar a publicar una nota negativa.

Y, también hay opiniones sobre que la propia falta de calidad de los informes limita el hecho de publicarlos:

Para que una evaluación sea útil para otras organizaciones ajenas a la intervención evaluada ha de ser bueno, conciso, constructivo, original, y está claro que normalmente no es así.

2.7. Análisis por correlaciones

Para un análisis de mayor profundidad pero sin pretensión de abordar causalidades, elaboramos cinco nuevas variables ficticias a partir de la suma simple de las frecuencias obtenidas en los cuestionarios. Las nuevas variables y su recorrido son:

- Participación de la ONGD: recorrido [0-9] por la suma de los nueve ítems en los que se señaló la participación a lo largo del ciclo de la evaluación.
- Participación del socio local: análoga a la anterior mediante la suma de frecuencias del socio local durante el proceso evaluativo.
- Suma de Utilidad: la suma de los tres ítems sobre la utilidad (proceso, conclusiones y recomendaciones); recorrido [3-12].
- Suma Equipo: suma de los nueve ítems que calificaban la acción del equipo evaluador; recorrido [9-36].
- Suma recursos: los tres ítems que preguntaron por tiempo, recursos humanos y dinero; recorrido [3-14].

Añadiendo a estas variables la del "Valor" global otorgada por la ONGD entre cero y diez podemos obtener información relevante para intentar determinar qué función explica la utilidad de la evaluación de los convenios y cuál no la explica.

Un primer acercamiento de las interrelaciones entre estas variables lo ofrece la matriz de correlaciones que se detalla en la Tabla 9.

TABLA 9. Matriz de correlaciones en torno a la utilidad del proceso evaluativo

	Valor	PART_ONGD	PART_SL	SUMA_equipo	SUMA_Util	SUMA_Recursos
Valor	1					
PART_ONGD	-0,030	1				
PART_SL	0,289	-0,005	1			
SUMA_equipo	0,923	0,098	0,228	1		
SUMA_Util	0,689	-0,078	0,243	0,708	1	
SUMA_Recursos	-0,197	0,078	-0,383	-0,115	-0,178	1

Elaboración propia.

La asociación estadística más fuerte se produce entre el *valor* que subjetivamente otorgan al proceso las ONGD y la percepción que tienen del trabajo de los evaluadores (0.92) y de la utilidad (0.69). También es alta la correlación entre la utilidad y el equipo (0,71). El resto de variables ofrecen niveles bajos de correlación. Esta correlación mostrada por el análisis cuantitativo, se fortalece con la información obtenida del análisis cualitativo en el que las ONGD muestran reiteradamente como una de las debilidades más importantes del proceso, (los cuellos de botella) a los propios equipos evaluadores, mencionando aspectos como: la falta de dominio de la metodología de evaluación y el uso de diversas técnicas de recogida de información, la escasez de tiempo dedicado en terreno a la recogida de información, falta de entendimiento entre los miembros de un mismo equipo, evaluadores no expertos en la temática o sector de evaluación, entre los principales.

Quizá pueda resultar de interés la correlación negativa entre la participación de los socios locales y los recursos invertidos (-0.38) pues más bien podría ser al revés: una mayor participación de los socios locales habría podido aumentar el tiempo, los RR.HH. y el coste monetario de la evaluación.

El *valor* atribuido al proceso de evaluación de los convenios no se explica, en el análisis cuantitativo, por la participación de la propia ONGD en el proceso (-0.03) ni por los recursos invertidos que incluso tienen una correlación levemente negativa (-0.19). Sin embargo, de la información cualitativa, sí es importante resaltar la importancia que las ONGD dieron a todo el proceso participativo para que la evaluación fuera realmente útil. De hecho, en el grupo de discusión hubo unanimidad en cuanto a la que la participación de la ONGD desde el principio del proceso era clave para evitar cuellos de botella. Hubo un comentario que resumiría esta percepción: *Si hay participación y un buen equipo de evaluación, todo lo demás viene dado.*

La *utilidad* ("suma_útil") percibida por las ONGD no aparece muy asociada a la falta de participación del socio local (0.24), ni a los recursos invertidos (-0.17).

Si, a modo de intento de modelización, regresamos la utilidad sobre el resto de las variables ficticias obtenemos los siguientes resultados (Tabla 10) a partir de la ecuación:

$$\text{Suma_Util}_i = \alpha + \beta \text{Valor}_i + \chi \text{Part_ONGD}_i + \delta \text{Part_SL}_i + \varepsilon \text{Suma_equipo}_i + \varphi \text{Suma_recursos}_i + \gamma$$

Ni el *valor* subjetivamente asignado, ni la *participación de los socios locales* aparecen estadísticamente asociadas a la utilidad. El único factor explicativo

significativo al 99% de confianza es el *equipo* evaluador. Para conocer de forma más detallada cuál de las nueve variables que componen la "Suma_equipo" se procedió a hacer la regresión de ésta sobre los nueve componentes. La única que resultó estadísticamente significativa fue "*equipo*"⁸. Lo que de nuevo claramente se sustentan, con todo el análisis cualitativo de la información, en donde las ONGD han puesto el peso de la utilidad de la evaluación, es decir, en la composición de buenos equipos evaluadores.

TABLA 10. Regresiones sobre la utilidad del proceso evaluativo

	1	2	3
Constante	4,07	4,02	4,18
	0,067	0,061	0,043
Valor	0,08		
	0,873		
Part_ONGD	-0,13	-0,13	-0,137
	0,404	0,342	0,334
Part_SL	0,03	0,03	
	0,749	0,723	
Suma_equipo	0,26	0,291***	0,295***
	0,14	0,000	0,000
Suma_recursos	-0,04	-0,05	-0,06
	0,720	0,691	0,563
R2	0,5339	0,5333	0,530
F	4,811**	6,286**	8,666**
Observaciones	27	27	27

Nota: Var dependiente: SUMA_Util; en cursiva los p-values. *** significativa al 99% y ** al 95%.

Elaboración propia.

⁸ Se recuerda que la pregunta a responder era "¿Cómo calificaría al equipo evaluador?". El coeficiente fue 1.80, error estándar 0.37 y estadístico t = 4.84. R² = 0.473.

Resultados muy similares se obtienen considerando el **valor** subjetivo del ejercicio evaluativo como variable dependiente. Como muestra la Tabla 11, únicamente la suma de las frecuencias otorgadas a los ítems que median al equipo evaluador aparece como estadísticamente explicativa del valor percibido.

TABLA 11. Regresiones sobre el valor de las evaluaciones realizadas

	1	2	3
Constante	-0,74	-0,68	-0,55
	0,457	0,449	0,526
Suma utilidad	0,015		
	0,873		
Part_ONGD	-0,09	-0,094	-0,09
	0,169	0,142	0,138
Part_SL	0,02	0,029	
	0,519	0,5	
Suma_equipo	0,317***	0,322***	0,326***
	0,000	0,000	0,000
Suma_recursos	-0,04	-0,04	-0,05
	0,48	0,459	0,285
R2	0,8762	0,8761	0,8734
F	29,75**	38,9**	52,93**
Observaciones	27	27	27

Nota: Var dependiente: Valor; en cursiva los p-values. *** significativa al 99% y ** al 95%.

Elaboración propia.

Análogamente a como se hizo para la utilidad, se calculó la regresión del valor sobre los 9 ítems de "suma_equipo"⁹. Esta vez los componentes significativos fueron

⁹ Los resultados no se muestran por espacio. Quedan disponibles para quien desee consultarlos.

“ajuste” entre el trabajo finalmente hecho por el evaluador y la oferta inicial, el equipo y la comunicación con la ONGD.

TABLA 12. Regresión del “valor” sobre los ítems del trabajo del equipo evaluador

		Parte del R ²
constante	0,21	0,0062
	<i>0,703</i>	
comunicación	0,33*	0,1162
	<i>0,088</i>	
equipo	0,84**	0,3267
	<i>0,002</i>	
ajuste	1,07***	0,4424
	<i>0,000</i>	
R2	0,8618	
F	49,91**	
Observaciones	28	

Nota: Var dependiente: Valor; en cursiva los p-values. *** significativa al 99%, ** al 95% y * al 90%.

Elaboración propia.

En resumen, según la experiencia concreta tenida en la evaluación de los convenios 2006–2007, y de acuerdo a las valoraciones otorgadas por 27 ONGD, el proceso ha resultado útil gracias al desempeño de los equipos evaluadores. Esta utilidad no está estadísticamente asociada al valor percibido por las ONGD, ni a la participación de los socios locales, ni a la ONGD, ni a los recursos invertidos.

Por último, llama la atención la falta de significatividad de la participación del socio local. En teoría las ONGD son defensoras de un modelo de cooperación inclusivo y participativo, pero en las evaluaciones aún no se ha traducido en una práctica común. Por ejemplo, dos entidades afirmaron que los socios locales no habían participado en ninguna de las fases de la evaluación y otra que participaron sólo en los TdR. Este dato puede dar lugar a una recomendación para incrementar en

el futuro este bajo nivel de involucración, ya que –al menos en teoría– quienes ponen en práctica las principales recomendaciones de una evaluación intermedia de un convenio, son los socios locales y los beneficiarios.

3. Conclusiones

En esta sección se recogen de forma sumaria las principales conclusiones obtenidas, ordenadas en función de los grupos de variables explicativas consideradas.

Respecto al momento

La mayoría de las ONGD considera muy positivo que haya habido un impulso para realizar evaluaciones intermedias. Una externalidad positiva fruto de esta iniciativa es que algunas de éstas Organizaciones están ahora llevando a cabo evaluaciones finales por propia iniciativa y con autofinanciación. Por otra parte, las ONGD demandan un mayor espacio para poder realizar las evaluaciones en el momento que consideren que puedan ser de más utilidad para la propia intervención y para la ONGD. El criterio para determinar ese mejor momento estará en función de la duración total de la intervención prevista inicialmente. Hay convenios que están pensados como una primera fase de un proceso de más largo plazo. En estos casos una evaluación al final del primer convenio (en vez de la intermedia) puede ser de mayor utilidad.

Motivo principal

La mayoría de las ONGD ha visto en las evaluaciones una posibilidad de aprendizaje y de mejora más allá de la obligatoriedad de la norma. Sin embargo, los intereses evaluativos de la ONGD no siempre se correspondían con los de la AECID. Algunas ONGD han vivido los intereses de la AECID como “imposición” lo que ha restado utilidad para las Organizaciones.

La AECID ha tratado de que sus intereses se vieran reflejados en las evaluaciones a través de su participación en la definición de los TdR. La apreciación de las ONGD es que la participación de la AECID en esta fase ha generado procesos demasiado largos, generado poco valor añadido y producido cierto “desgaste” en la mayoría de las Organizaciones.

Externalización y elección de la empresa evaluadora

Para las ONGD, la función principal de la evaluación del convenio era el aprendizaje y no la rendición de cuentas. Es por ello que consideran que una evaluación mixta hubiera permitido un mayor aprendizaje y nivel de apropiación. Los equipos mixtos pueden generar sinergias ya que la ONGD y su SL pueden aportar de forma más eficiente a los evaluadores externos el conocimiento sobre el país, sobre el contexto de ejecución y sobre el sector de la intervención. Por su parte, los evaluadores externos pueden aportar mayor conocimiento sobre el proceso evaluativo (metodología y herramientas) y contraste con otras experiencias comparables que quizá la ONGD y el SL no tengan. Quien finalmente pone en práctica la mayoría de las recomendaciones es el SL con el apoyo de la ONGD, con lo que si su participación se traduce en mayor motivación y conocimiento fundado de las conclusiones que originan dichas recomendaciones, su ejecución real y transformativa puede ser mayor, y la evaluación puede ser más influyente.

Exigencia de evaluadores locales

En convenios multi-país, este requisito ha generado los frutos esperados ya que generalmente se ha perdido la mirada regional del convenio. El supuesto de que los locales pueden tener valor añadido en cuanto a conocer mejor el contexto regional, no se ajusta a la realidad. Un evaluador de origen de uno de los países del convenio no siempre posee un conocimiento en profundidad del resto de los países del convenio.

Otra dificultad encontrada ha sido la falta de conocimiento del evaluador local del sistema de la cooperación española, dando origen a recomendaciones poco prácticas o contrarias a la normativa española vigente.

Participación de las ONGD

Tanto entre las organizaciones con mayor cultura evaluativa como entre las ONGD que empiezan a interesarse por la evaluación, la participación aparece como un elemento clave en su discurso, que es corroborado con los datos del análisis cuantitativo. Destacan sobre todo la necesidad de un alto nivel de participación en las primeras fases del proceso a la hora de definir las preguntas de evaluación y los TdR. Debido a que eran evaluaciones externas, el nivel de participación es menor en las fases de recogida e interpretación de la información así como en la elabo-

ración de conclusiones y de recomendaciones. Sin embargo, en esta última fase, las ONGD reconocen que depende de la flexibilidad del propio equipo evaluador el que se tengan en mayor o menor consideración las valoraciones y opiniones de las ONGD, convirtiéndose en algunos casos en una ardua negociación ante la diversidad de interpretaciones.

Participación de los Socios Locales

La participación de los socios locales es relativamente baja. En opinión de las ONGD, esta menor participación se debe a la falta de interés porque la evaluación no se concibe como una oportunidad de mejora sino como un trámite más, una imposición, debido a la falta de cultura evaluativa en sus organizaciones.

Sobre la calidad del trabajo hecho por los evaluadores

El trabajo realizado por los evaluadores, en términos generales, es calificado de bueno por las ONGD (aunque la cuarta parte de las ONGD califican este trabajo como regular). En general, la valoración global sobre el trabajo realizado se calificaría de notable (promedio de 6,5 y moda de 7) en una escala del 1-10.

La menor satisfacción de las ONGD sobre el trabajo de los evaluadores se muestra en la fase de recogida de información y en la que la evaluación haya sido completa. Se aprecia una tendencia a recoger y resaltar lo negativo y no identificar o señalar también las fortalezas de las intervenciones. La mayor satisfacción se muestra en la comunicación mantenida entre los equipos evaluadores y las ONGD.

La información recogida cualitativamente se centró sobre todo en la preocupación que las ONGD mostraron en cuanto a los propios equipos de evaluación, las principales preocupaciones identificadas son:

- Contratar a una empresa determinada no es garantía de buen trabajo ya que a veces éstas suelen subcontratar estos trabajos. En definitiva, la calidad del equipo evaluador lo aportan las propias personas que lo componen y todavía no identifican empresas con suficiente marca que las garantice un buen producto evaluativo.
- Otro tema importante en cuanto a la composición del equipo evaluador, es el dilema de priorizar expertos en evaluación o expertos en el sector que se

evalúa. La experiencia tenida en estos convenios apunta a la priorización de expertos en evaluación tratando de compensar las limitaciones en el conocimiento específico de la temática por otras vías.

- La principal debilidad que apuntaron las ONGD en cuanto a la profesionalidad del equipo evaluador era la dificultad de encontrar verdaderos expertos en evaluación con amplios conocimientos metodológicos.

Sobre la utilidad

La percepción revelada sobre la utilidad del proceso, las conclusiones y recomendaciones parece ser alta o muy alta en la mayoría de las ONGD. Esta valoración conduce a querer conocer *las consecuencias* concretas que ha tenido la evaluación sobre las intervenciones de las ONGD. La primera consecuencia es contar con un plan de mejora. La mayoría de las ONGD afirma que ya cuenta con él o están en proceso de elaboración. Muchas de las ONGD afirman que su plan de mejora es más completo y de mayor alcance que la agenda firmada con la AECID. Además, el 93% de las ONGD afirman que se han obtenido recomendaciones útiles para ser aplicadas a otras intervenciones de la organización (replicabilidad) además de a la continuidad de los convenios que tuvieron una segunda fase.

En cuanto al agente para el que la evaluación ha sido más útil, las ONGD perciben que ellas mismas son el principal beneficiario de la evaluación, seguidos de sus socios locales, quedando en último término la propia AECID y los beneficiarios, aunque la mayoría reconoce que de manera indirecta, en la medida en que se mejoran los procesos, la evaluación también resulta útil para los beneficiarios últimos. Las ONGD perciben que la AECID no tiene capacidad para asimilar todas las evaluaciones realizadas y por tanto esto restaría utilidad de las mismas para este agente.

Sobre los recursos utilizados

No se deduce del análisis cuantitativo una falta de recursos (económicos, humanos y de tiempo) que se correlacione significativamente con la utilidad-credibilidad de la evaluación. Sin embargo, desde un punto de vista cualitativo surgen reflexiones interesantes sobre los condicionantes de los recursos para obtener trabajos de calidad. Un testimonio afirmó que "el dinero para hacer la evaluación parece determinar la composición del equipo evaluador, y el tiempo que se emplea para

recoger información en terreno. Estas son dos cuestiones básicas para hacer un trabajo de calidad”.

Se detectaron algunas quejas por parte de las ONGD relativas a la escasez de tiempo dedicado al trabajo de campo por parte de algunos evaluadores lo que limitó el alcance de la recogida de información, así como deficiencias importantes en la composición de los equipos evaluadores.

También existe la percepción generalizada de que el tiempo es escaso no para realizar la evaluación en sí, sino para procesar el aprendizaje obtenido y para implementar mejoras al quedar poco tiempo de vida del convenio una vez obtenido el informe de evaluación.

Por último, también se manifiesta de forma generalizada que el tiempo empleado por el personal de la ONGD ha llegado a ser excesivo sobre todo durante la negociación de los TdR con AECID.

Limitaciones, cuellos de botella

Los principales cuellos de botella recogidos por las ONGD que restan utilidad a las evaluaciones realizadas son los siguientes (se recogen solo aquellos que más se repetían en la información cualitativa):

- *Metodología inadecuada* para convenios regionales; se pierde la visión de región y se dificulta tener que evaluar en todos los países con la misma contratación de equipo de evaluación.
- Limitaciones en el *equipo evaluador*: Falta de dominio del idioma local, falta de coordinación entre los miembros del equipo, escasa formación en metodologías y herramientas de recogida de información de evaluación (a menudo no debidamente justificadas).
- *Objeto* de la evaluación *demasiado amplio*: El hecho de que se obligue a evaluar bajo los criterios del CAD no permite dar tanta importancia a los procesos que en una evaluación intermedia debería ser el elemento central.

Difusión

Los datos muestran que sólo 6 ONGD (21%) han respondido que han difundido de alguna forma su evaluación. El principal motivo de la falta de publicación parece ser la falta de cultura evaluativa (a mayor cultura evaluativa, mayor difusión de los resultados de las evaluaciones o de las propias evaluaciones). De hecho muchas ONGD contestaron que éste era un tema que aún no se había debatido o reflexionado suficientemente en el interior de su organización. Aunque se recogen otras opiniones como la falta de calidad de los informes de evaluación y el temor de la publicación de malos resultados, estas opiniones son minoritarias.

Factores determinantes de la utilidad

De las variables estudiadas en esta investigación como endógenas a la utilidad de la evaluación, tanto el análisis cuantitativo como cualitativo revela que el propio equipo de evaluadores es el mayor condicionante para que la utilidad sea alta para las ONGD, siendo un aspecto muy importante el ajuste entre el trabajo finalmente hecho y la oferta inicial realizada por los evaluadores. El análisis cualitativo, también refleja como significativa para la utilidad de la evaluación, la amplitud y profundidad de la participación de la propia ONGD en todo el proceso de evaluación.

4. Referencias

AECID (2005) Orden AECI/1303/2005, de 27 de abril, por la que se regulan las bases para la concesión de subvenciones a organizaciones no gubernamentales de desarrollo, para la realización de intervenciones en materia de cooperación internacional para el desarrollo. BOE 113 de 12 de mayo.

– (2009a) Guía para evaluaciones de proyectos y convenios de ONGD. http://www.aecid.es/galerias/ongds/descargas/Guxa_evaluacixn_sept09def.pdf

– (2009b) Normas de seguimiento y justificación de proyectos y convenios de cooperación al desarrollo subvencionados a las organizaciones no gubernamentales de desarrollo. BOE 93 de 16 de abril.

LARRÚ, J.M. (2002) “Razones para implantar un sistema de evaluación en las ONGD y sus consecuencias”, *Revista de Fomento Social* 57, 683–710.

– (2005) “La Declaración de París–2005: Principios para una ayuda eficaz y aplicaciones para las ONGD”, *Revista de Fomento Social* 60, 243–281.

– (2009) *La Ayuda al Desarrollo, ¿Reduce la Pobreza? Eficacia y evaluación en la cooperación para el desarrollo*, Madrid, Biblioteca Nueva.

OECD (2002) *Glosario de los principales términos sobre evaluación y gestión basada en resultados*. Evaluation and Aid Effectiveness N°6. OECD–DAC. Paris. <http://webnet1.oecd.org/pdf/M00030000/M00030075.pdf>

OECD–DAC (2005) “The Paris Declaration on Aid Effectiveness”. 2nd High Level Forum on Aid Effectiveness. Development Co–operation Directorate. DCD–DAC. Feb 28–March, 2.

– (2008) “The Accra Agenda for Action”. 3rd High Level Forum on Aid Effectiveness. Development Co–operation Directorate. DCD–DAC. Sep, 2–4.

– (2011) “Busan Partnership for Effective Development Co–Operation”, DCD–DAC. Nov 29–Dec, 1.

STAKE, R.E. (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona, Graó.