



**Prácticas recomendadas
para las evaluaciones periódicas**

Guía para planificadores de evaluación y evaluadores

Versión final

Secretaría del PEFA
1 de febrero de 2010

Resumen

Desde su creación, en 2005, el Marco para la Medición del Desempeño del Programa de Gasto Público y Rendición de Cuentas Financieras (PEFA) se ha utilizado en más de 150 evaluaciones del sistema de gestión de las finanzas públicas (GFP) de países. Aunque muchas evaluaciones recientes son “periódicas”, a menudo no satisfacen uno de los objetivos fundamentales del marco: proporcionar una imagen clara de los cambios concretos que se han producido en el desempeño desde la evaluación inicial.

Por consiguiente, se ha pedido a la Secretaría que formule orientaciones sobre “prácticas recomendadas” y en esta nota se formulan recomendaciones para cada etapa de una evaluación periódica típica.

En resumen, las orientaciones son las siguientes:

PARA GESTORES DE EVALUACIÓN/PATROCINADORES

NOTA CONCEPTUAL/TÉRMINOS DE REFERENCIA

- No realizar evaluaciones con demasiada frecuencia: el cambio requiere tiempo, y el intervalo recomendado para repetir la utilización del marco del PEFA **es de al menos tres años**.
- Asegurarse de que **la nota conceptual/los términos de referencia sean lo suficientemente minuciosos**, y de que todas las partes interesadas los entiendan claramente.
- **Facilitar el acceso a la documentación** de la evaluación anterior, ya sea que esta se haya publicado o no, y a las notas que no se han publicado y, de ser necesario, a los comentarios verbales del jefe de equipo de la evaluación anterior.
- El **organismo principal que se ocupó de la evaluación anterior** (si no es el mismo que se ocupa de la evaluación periódica) debería formar parte del grupo de referencia para la evaluación reiterada.
- Utilizar **el mismo equipo que antes constituiría una ventaja**, siempre que el equipo hubiese elaborado un informe de buena calidad y que entretanto sus miembros no se hayan dedicado a diseñar e implementar en el país de que se trate una reforma de la GFP y el fortalecimiento de la capacidad, ya que esto podría presentar un conflicto de intereses con la evaluación.

PARA EVALUADORES

PREPARACIÓN

- **Comunicarse con el jefe de la evaluación anterior** a fin de obtener la información adicional que sea necesaria sobre los datos y las fuentes (incluidos los comentarios de otros expertos evaluadores).

- No encarar la evaluación **suponiendo que sus predecesores se equivocaron en sus calificaciones ni que estas eran totalmente correctas.**
- **Utilizar la evaluación anterior como punto de partida** para explicar la trayectoria de cambios registrados en el sistema de GFP en materia de desempeño (antes que examinar por separado cada indicador de la evaluación anterior y luego comparar las calificaciones, ya que con ello se corre el riesgo de no identificar las diferencias en las definiciones, etc.).

TRABAJO EN EL TERRENO

- **Verificar la base sobre la que se asignó la calificación anterior** e identificar todos los indicadores sobre los que no se dispone de suficiente información para esta verificación.
- **Tomar notas completas y detalladas sobre los datos y las fuentes** correspondientes a cada calificación (aunque deban mantenerse en un archivo confidencial que no se incluirá en el informe propiamente dicho) para su seguimiento en el futuro.
- No utilizar la calificación “D” cuando no se dispone de información suficiente para asignar una calificación.
- **No tratar de volver a calificar la evaluación anterior:** esas calificaciones se han utilizado en documentos que, ya sea que se hayan publicado o no, habrán sido utilizados por diversas partes interesadas.

REDACCIÓN DEL INFORME

- Si se descubre que hubo errores en calificaciones anteriores, explicarlo en el análisis de los cambios en las calificaciones de los indicadores. **Puede insertarse una nota en el cuadro panorámico de los indicadores a fin de explicar que la calificación presente y la anterior no son equiparables.**
- **Ser precisos en relación con las fuentes de datos** (para ayudar a los futuros evaluadores).
- **Incluir la nota conceptual/los términos de referencia como anexo** al informe.
- **En la evaluación resumida:**
 - **Presentar una visión general de los cambios que se han producido en materia de calificación del desempeño** desde el informe anterior y un cuadro con las calificaciones para N y después para N+1.
 - **Hacer una lista de los indicadores** que han empeorado, los que han mejorado y los que se mantienen sin cambios.
 - **No realizar la agregación de resultados de “n” a “n+1” para señalar una mejora porcentual.**
- **En la descripción:**
 - Proporcionar **una visión general de los problemas que se plantean con las viejas calificaciones** como consecuencia de cambios en los datos y las definiciones.

- **Identificar el impacto de las discrepancias** con las calificaciones de la primera evaluación.
- **Identificar los motivos por los que se han producido cambios en las calificaciones de las dos evaluaciones, indicador por indicador**, a fin de que el lector pueda comprender exactamente lo que ha ocurrido.

1 Objetivo de esta nota

Desde 2005, el Marco para la Medición del Desempeño del PEFA se ha utilizado para realizar evaluaciones de los sistemas de GFP en más de 100 países. Hasta ahora, la mayor parte de las evaluaciones que se han realizado han sido evaluaciones básicas, pero, tal como es de esperar, las evaluaciones reiteradas son cada vez más frecuentes, debido a que en algunas regiones la cobertura básica ya casi se ha completado y han transcurrido casi tres años desde que se realizó la primera evaluación.

Aunque una proporción creciente de evaluaciones son reiteradas, a menudo estas no satisfacen uno de los propósitos fundamentales del marco: brindar una imagen clara de los cambios concretos que se han producido en el funcionamiento del sistema desde la evaluación inicial. En algunos casos, los progresos se describen en la evaluación resumida, pero no se hace un seguimiento minucioso de estos a nivel de los distintos indicadores individuales.

Como cada vez se realizan más evaluaciones reiteradas, los evaluadores, los funcionarios de los Gobiernos y los representantes de organismos de desarrollo han pedido a la Secretaría que formule orientaciones sobre “prácticas recomendadas” para la realización de evaluaciones periódicas, lo que ha motivado la preparación de esta nota.

2 Contexto: Utilización del marco

El Marco para la Medición del Desempeño en materia de GFP es uno de los tres componentes de la “estrategia reforzada” de apoyo a las reformas de la GFP. Estos componentes son: i) una estrategia y plan de acción para la reforma de la GFP impulsados por los países, ii) un programa de trabajo sobre la GFP coordinado y plurianual de donantes que preste apoyo a la estrategia de reforma de la GFP de los Gobiernos y esté en armonía con ella y, iii) una base común de información.

El marco es un instrumento para llevar a la práctica el tercero de los componentes, una base común de información, y, en particular, se ha diseñado para medir el desempeño en materia de GFP a lo largo del tiempo en el caso de países que se encuentran en diferentes niveles de desarrollo, al permitir establecer un punto de referencia para evaluar los progresos alcanzados.

Es de esperar que la aplicación reiterada y coherente del marco permita a las partes interesadas determinar hasta qué punto dicho marco mejora o no el desempeño de la GFP. Además, en el informe de desempeño en materia de GFP se reconocerán los esfuerzos realizados por cada Gobierno para reformar su sistema de GFP al describirse las medidas de reforma recientes y en curso, aunque estas tal vez no hayan influido aún en el desempeño relacionado con la GFP.

3 Coherencia

Es evidente que para poder comparar dos evaluaciones que se han llevado a cabo con cierto tiempo de diferencia, en ellas debe haberse aplicado la metodología en forma coherente. Suponiendo que en la evaluación inicial se aplicó rigurosamente la metodología, en teoría las partes interesadas querrán saber si se ha hecho un seguimiento preciso de los progresos a lo largo del tiempo mediante el uso de:

- los MISMOS asesores;
- el MISMO patrocinador;
- las MISMAS contrapartes;
- las MISMAS fuentes de datos.

Sin embargo, es poco probable que todos los elementos anteriores, o que quizá incluso alguno de ellos, estén disponibles para reiterar la evaluación. Por ejemplo, si para la evaluación anterior se utilizaron consultores, resulta poco probable que tres años después se pueda utilizar el mismo equipo de consultores (debido a la naturaleza de los cambios en los patrones de trabajo de estos profesionales). Algunos funcionarios gubernamentales habrán cambiado de puesto con el transcurso de los años. Además, menos de la mitad de las reiteraciones han sido patrocinadas por el mismo organismo de desarrollo, y algunas fuentes de datos, como los comentarios de los patrocinadores y de expertos evaluadores (incluida la Secretaría) sobre los proyectos del informe de la evaluación anterior, pueden infringir las normas de confidencialidad.

También existe la posibilidad de que la evaluación anterior no se haya considerado de buena calidad, en cuyo caso se plantea otro interrogante: ¿ofrece esta evaluación el punto de referencia confiable tan necesario para realizar un seguimiento significativo de los progresos alcanzados? Y de no ser así, ¿cómo deberían proceder los evaluadores encargados de la evaluación reiterada?

Habida cuenta de estas cuestiones de índole práctica, se torna crucial la cuestión de determinar qué es una “práctica recomendada” para los evaluadores cuando llevan a cabo una evaluación reiterada.

4 Temas señalados en las evaluaciones reiteradas

El principal objetivo de una evaluación reiterada es seguir el desempeño desde la evaluación anterior. Para ello, es necesario garantizar que se comparen elementos equivalentes. Esto significa entender los datos que sirvieron de base para la asignación de las calificaciones en la evaluación anterior. Es posible que esto no siempre sea obvio: por ejemplo, las cifras relativas a los gastos e ingresos de las que disponían las personas que realizaron la evaluación anterior pueden haber cambiado en virtud de la verificación realizada por auditores, y, por consiguiente, los evaluadores encargados de la reiteración de la evaluación tal vez reciban información fiscal diferente correspondiente al mismo año. A continuación se analizan en más detalle estas y otras cuestiones.

Cuando se examinan las más de 20 evaluaciones periódicas realizadas hasta ahora, se observan claramente dos cosas. En primer lugar, aunque en general se entiende que las medidas de reforma suelen tardar en ejercer su impacto, casi todas esas evaluaciones se terminaron en un tiempo más reducido que el período de tiempo recomendado en el marco del PEFA (de tres a cinco años). En segundo lugar, la calidad de esas evaluaciones reiteradas no es uniforme; algunas son de buena calidad, pero otras no lo son. Son varios los factores que podrían explicar estas deficiencias:

- En primer lugar, aunque el equipo de evaluación sea consciente de que está reiterando una evaluación, sus miembros pueden haber tomado deliberadamente la decisión de no consultar el informe anterior, como surge de esta cita sacada de un informe de evaluación: *“El equipo quería alcanzar conclusiones independientes sin verse influido por evaluaciones anteriores. Esto se discutió con las partes interesadas durante la evaluación”*. En esta situación, se torna mucho más difícil explicar las razones de los cambios en las calificaciones.
- En segundo lugar, es posible que en la nota conceptual/los términos de referencia no se haya especificado en suficiente detalle qué se esperaba que hicieran los evaluadores para hacer el seguimiento de los progresos alcanzados.
- En tercer lugar, los evaluadores tal vez no hayan visto la evaluación anterior y, si esta fue patrocinada por un organismo principal diferente, tal vez no hayan tenido acceso a los documentos preliminares o ni siquiera al informe final en caso de que no se hubiere dado acceso público a él; por lo tanto, sienten que no están en condiciones de hacer una comparación.
- En cuarto lugar, es posible que los equipos de evaluación no tengan acceso a los comentarios de expertos evaluadores (incluida la Secretaría) sobre el *proyecto* de informe de la evaluación anterior, en el que pueden haberse señalado cuestiones relativas a la calificación que **no** se tomaron en cuenta en la versión *final* del informe, y, por lo tanto, tal vez no dispongan de toda la información pertinente necesaria para procurar hacer una comparación.
- En quinto lugar, los equipos de evaluación pueden tener un punto de vista diferente sobre lo adecuada que resulta la calificación de la evaluación anterior, y, por consiguiente, quizá no puedan realizar una comparación significativa.
- Por último, algunas evaluaciones reiteradas constituyen un intento por establecer un punto de referencia más preciso que el de la evaluación anterior, en cuyo caso puede considerarse que no tiene sentido realizar una comparación.

5 Ejemplos de prácticas recomendadas

La Secretaría definiría como satisfactoria a una evaluación reiterada en la que se verifica la base sobre la que se asignó la calificación anterior y donde se identifican todos los indicadores sobre los que no se dispuso de información suficiente para realizar esa verificación. Además, una evaluación reiterada satisfactoria debería tener en cuenta todos los errores obvios en el uso de datos para otorgar la calificación anterior a fin de poder comparar elementos equivalentes.

Algunos ejemplos de deficiencias comunes señaladas en los comentarios de la Secretaría que se han extraído de las revisiones de informes preliminares recientes son:

- *“No se hace hincapié en el seguimiento de los progresos alcanzados. En el informe no se mencionan los factores que explican los progresos por indicadores, además de proporcionarse las calificaciones comparadas”.*
- *“La falta de comparación de las calificaciones de 200X y 200Y por indicador (y posiblemente subindicador) no permite comprender los cambios (y la falta de cambios)”.*
- *“Sólo se presenta una comparación o explicación para algunos indicadores”.*
- *“No hay ningún cuadro panorámico en el que se muestren las calificaciones anteriores y actuales, junto con explicaciones acerca de los progresos alcanzados o la falta de progresos”.*

Los tipos de cuestiones enumeradas antes se abordarán si:

- en la evaluación resumida se presenta una breve reseña de los cambios de las calificaciones del desempeño desde el informe anterior y un cuadro en el que consten los dos conjuntos de calificaciones;
- se mencionan los motivos de los cambios en las calificaciones entre las dos evaluaciones, por indicadores;
- se incluye como anexo un panorama minucioso de los progresos alcanzados entre las dos evaluaciones.

Las calificaciones de los indicadores proporcionarán una visión general rudimentaria de los cambios que se han producido en el transcurso del tiempo, pero los distintos aspectos pueden variar en forma diferente, y el desempeño quizá no cambie lo suficiente como para hacer cambiar la calificación del indicador (en cuyo caso puede resultar apropiado usar una flecha). Por lo tanto, hará falta una explicación más detallada.

A fin de ilustrar las prácticas recomendadas, la Secretaría presenta los siguientes ejemplos extraídos de informes recientes:

- En la evaluación resumida:

..... *“En el cuadro A1, siete indicadores parecen haber empeorado desde 2005. De hecho, algunas de las calificaciones de 2005 ahora parecen exageradas, como se explica en los cuadros comparativos de todo el informe. Solo hay un indicador, ID-1, que parece haber empeorado realmente. Ello es sorprendente, habida cuenta de*

las plantillas elaboradas en el marco del programa de reforma, y tiene que ser objeto de seguimiento en el contexto del estudio. Al parecer, no hubo modificaciones en 12 indicadores, aunque algunos (ID-10, 15, 18, 26 y 27) registraron cambios significativos, que no han sido suficientes para influir en sus calificaciones. Otros 12 indicadores parecen haber mejorado, aunque una nueva calificación de ID-5 en 200X los reduce a 11 (ID-4, 6, 9, 13, 14, 17, 20, 21, 22, 24 y D-2). Algunos de estos reflejan la aplicación y el aumento del uso de los nuevos sistemas de gestión financiera y otras reformas. Sin embargo, quedan por abordar de manera más eficaz algunos puntos débiles, como se señala en el resumen”.

Cuadro A1			
		Calificaciones de 200X	Calificaciones de 200Y
ID-1	<i>Gasto agregado en comparación con el presupuesto original aprobado</i>	B	B
ID-2	<i>Composición del gasto en comparación con el presupuesto original aprobado</i>	C	C
ID-3	<i>Ingresos agregados en comparación con el presupuesto original aprobado</i>	A	A
ID-4	<i>Saldo y seguimiento de los atrasos de pagos de gastos</i>	D	D+

- Para los distintos indicadores:

“Comparación entre 2004 y 2006. A partir de la información disponible en 2007, señalamos que la calificación del indicador ID-1 en 2004 en verdad estaba equivocada, en particular debido a la inclusión, dentro del cálculo, de los gastos del servicio de la deuda. En ese entonces, las desviaciones de los gastos registradas fueron de +3,7% en 2002, -10,8 % en 2003 (debido a las inundaciones que asolaron el país en 2000 y 2001, que hicieron necesario realizar grandes gastos en ayuda y reconstrucción que fueron difíciles de presupuestar en forma precisa) y -3,7% en 2004. Nuestro análisis actualizado de los datos correspondientes al año 2004 (sobre la base del estado final de las cuentas) revela una desviación de -5,19% en el ejercicio de 2004, que habría redundado en una calificación B”.

- Además, un anexo en el que conste un panorama detallado en el que se presente la calificación PEFA N y PEFA N+1 para cada indicador y una descripción de los progresos entre N y N+1:

		Calificaciones de 200X	Calificaciones de 200Y	Cambios entre 0X y 0Y

ID-1	Gasto agregado en comparación con el presupuesto original aprobado	C	B	Centrarse constantemente en mantener la disciplina fiscal ha redundado en el aumento de la previsibilidad del presupuesto general.
ID-2	Composición del gasto en comparación con el presupuesto original aprobado	D	D	Se ha producido una ligera mejora en la varianza media del gasto de los MDO por encima del total de desviaciones del gasto. Esto puede deberse tanto a un financiamiento sostenido de los programas de políticas prioritarias como a las mejoras en la liberación de recursos para los MDO durante todo el año. No obstante, la varianza presupuestaria sigue siendo alta.
ID-3	Desviaciones de los ingresos agregados en comparación con el presupuesto original aprobado	A	A	Las constantes proyecciones moderadas de ingresos y la mejora de las medidas de aplicación han hecho que la recaudación de ingresos se mantenga elevada en relación con las cantidades presupuestadas.
ID-4	Saldo y seguimiento de los atrasos de pagos de gastos	D+	B+	
i)	<i>Saldo de los atrasos de pagos de gastos y cambio reciente en el saldo</i>	<i>D</i>	<i>B</i>	La estrategia quinquenal de los Gobiernos para reducir los atrasos de pagos de gastos ha generado reducciones interanuales de los atrasos generales.
ii)	<i>Disponibilidad de datos para hacer un seguimiento del saldo de los atrasos de pagos de gastos</i>	<i>B</i>	<i>A</i>	Los informes trimestrales periódicos ahora son compilados por el Interventor de la Auditoría Interna

6 Diferencias con calificaciones anteriores

A fin de poder seguir y entender debidamente los cambios en el desempeño en materia de GFP y poder confiar en la comparación de las calificaciones de un indicador a lo largo del tiempo, es necesario que el texto del informe de desempeño en materia de GFP permita a los lectores entender qué y cuánto ha cambiado.

También debe recordarse que no todos los cambios en las calificaciones representan siempre un cambio en el funcionamiento del sistema de GFP. A continuación se presentan posibles razones de los cambios (extraídas de evaluaciones recientes):

- **Diferencias en el acceso a la información**

Ejemplo del ID-7: *“En esta ocasión se ha tomado en cuenta un estudio de las operaciones cuasifiscales de las empresas paraestatales. En la evaluación anterior no se disponía de un estudio de este tipo. Por consiguiente, la calificación es más baja esta vez, pero no hay indicios de que nada haya empeorado”*.

Otro ejemplo: ID-4: *“Esta vez la evaluación se basa en una definición diferente de atrasos. La última vez la evaluación se basó en la definición de contabilidad y, por lo tanto, solo incluía pagos que se habían aceptado y procesado para que los pagara la Tesorería. De esta forma, el volumen de los atrasos ha aumentado, pero es posible que el desempeño no haya empeorado”*.

- **Diferencias en el muestreo**

Ejemplo del ID-21: *“En un informe anterior, la evaluación se basaba en la información sobre la situación de los Ministerios de Agricultura, Justicia y Finanzas. En esta ocasión se basa en información de los Ministerios de Educación y Justicia. Por lo tanto, no existe la seguridad de que una comparación directa resulte válida”*.

- **Diferencias en la interpretación**

Ejemplo del ID-24 ii): *“El año pasado, la emisión de los informes de ejecución del presupuesto se llevó a cabo de 4 a 7 semanas después de finalizado el trimestre, con un promedio de 5 semanas y media de retraso, y se asignó una calificación B. En la última evaluación, el desempeño fue exactamente el mismo, pero se otorgó una calificación C, debido a que dos informes se retrasaron más de seis semanas”*.

- **Metodología para asignar calificaciones (se producen errores)**

Debe explicarse la influencia que tienen los errores en los cambios en las calificaciones. También se presentan algunos ejemplos:

- *“otorgar una calificación errónea a pesar de que las pruebas son muy claras”*;
- *“utilizar la calificación M1 de agregación de dimensiones cuando se debería haber utilizado el M2”*;
- *“otorgar ‘+’ al indicador de una sola dimensión”*.

La Secretaría **recomienda** que, si los encargados de las evaluaciones reiteradas se encuentran con alguna de las situaciones señaladas antes, **no deberían procurar realizar una recalificación expresa de la evaluación anterior**, ya que este trabajo puede haber sido utilizado y ello podría dar lugar a interminables debates. Si en el trabajo anterior se produjeron errores irrefutables en las calificaciones, estos pueden explicarse en el análisis de los cambios en las calificaciones de los indicadores.

Puede insertarse una nota en el cuadro panorámico de indicadores para explicar que las calificaciones actuales y previas no son equiparables, o cómo una opinión diferente acerca del desempeño en la evaluación anterior puede haber influido en las conclusiones acerca de la dirección del cambio en el funcionamiento del sistema de GFP registrado desde entonces.

Asimismo, resulta importante señalar que la evaluación de cada indicador debe vincularse a la evaluación anterior, esto es, que la atención debe centrarse en explicar la trayectoria del desempeño utilizando como punto de partida la evaluación anterior, en lugar de realizar una evaluación del indicador separada de la evaluación anterior y después comparar las calificaciones. Si se adopta este último enfoque, se corre el riesgo de ignorar el tipo de diferencias antes mencionadas.

Por último, tal vez convenga decir por qué ha mejorado el desempeño, aunque ello no siempre es posible, debido a que puede requerir un análisis más detallado que el que se supone que debe proporcionar el informe de desempeño en materia de GFP, y que se busquen las causas básicas (como los factores relacionados con la capacidad).

7 Información sobre los diferentes indicadores

La Secretaría recomienda que esto se haga en un cuadro, como consta a continuación:

Indicador	Calificación de 2006	Calificación de 2010	Cambio en el desempeño	Otros factores
ID-1	C	B	El desempeño parece haber mejorado teniendo en cuenta lo siguiente: 2006: 6%, 11%, 18% 2010: 5%, 11%, 6%	No está claro si todos los fondos externos para proyectos se excluyeron de los datos para la evaluación de 2006, aunque probablemente se trate de una cuestión de escasa importancia.

ID-4 i)	A	C	El cambio en el desempeño es incierto, a pesar de que los atrasos aumentaron del 1% en 2006 a 6% en 2010.	En la evaluación de 2006 solo se utilizaron datos relativos a órdenes de pago pendientes y no a facturas vencidas.
---------	---	---	---	--

8 Conclusiones sobre los cambios generales en el funcionamiento del sistema de GFP

A medida que ha aumentado el número de evaluaciones reiteradas, los autores han procurado varias veces crear una medida agregada de los cambios en el desempeño. Tal medida agregada puede resultar especialmente interesante en los casos en que algunos indicadores dan muestras de progreso mientras que otros indican un empeoramiento del desempeño. La agregación comúnmente entraña tanto la conversión de las calificaciones ordinales de los indicadores en valores numéricos como la asignación de ponderaciones a los distintos indicadores. Sin embargo, debido a que no existe un método científicamente correcto para realizar estas conversiones, la Secretaría no apoya ni recomienda un enfoque en particular. Para un análisis pormenorizado de las cuestiones metodológicas que se plantean cuando se procura agregar las calificaciones del PEFA, véase la nota titulada *Cuestiones relativas a la comparación y la agregación de los resultados de las evaluaciones del PEFA en distintos países a lo largo del tiempo* en el sitio web del PEFA (www.pefa.org).

Como queda claro en la nota de orientaciones, el enfoque que se recomienda consiste en comparar las calificaciones correspondientes a cada indicador y utilizar la parte descriptiva del informe de desempeño en materia de GFP a fin de extraer conclusiones sobre los motivos de las diferencias en las calificaciones y, por consiguiente, cualquier cambio en el funcionamiento del sistema de GFP. La sección titulada “Evaluación resumida” de la evaluación debería utilizarse para extraer conclusiones generales sobre la dirección y medida de los cambios y, cuando corresponda, debería hacerse hincapié en la importancia relativa de diversos cambios en los indicadores y los vínculos entre ellos.

9 Nota conceptual/términos de referencia

El seguimiento de los progresos a lo largo del tiempo debería abordarse de manera cuidadosa y precisa en la nota conceptual/los términos de referencia. A fin de poder vigilar adecuadamente y entender los cambios en el desempeño en materia de GFP y poder confiar en la comparación de las calificaciones de un indicador a lo largo del tiempo, es necesario que la parte descriptiva del informe de desempeño de la GFP permita a los lectores entender qué y cuánto ha cambiado.

Por consiguiente, los evaluadores deben ser conscientes de que se espera que realicen una comparación de evaluaciones y un seguimiento del desempeño a lo largo del tiempo (proporcionando explicaciones). En la nota conceptual/los términos de referencia se debería exigir que el equipo explique todos los factores que repercuten en el cambio de calificación, indicador por indicador; identifique los cambios en el desempeño y garantice que cualquier lector pueda seguir los cambios que se han producido desde la evaluación anterior. Esto puede conformar una parte importante de la redacción del informe, aunque la existencia de una evaluación anterior (de buena calidad) en general represente una ventaja en el trabajo de evaluación.

Para que de la evaluación reiterada surja una trayectoria de progreso que resulte útil, es importante que la evaluación de cada indicador se vincule con la evaluación anterior, es decir que se centre la atención en explicar la trayectoria de desempeño utilizando como punto de partida la evaluación anterior, en lugar de realizar una evaluación del indicador separada de la evaluación anterior y después comparar las calificaciones (el peligro de este enfoque es que en la trayectoria pueden plantearse diferencias en el muestreo de definiciones, puede surgir nueva información de la que antes no se disponía o pueden variar las interpretaciones o los criterios en los casos límites). La nota conceptual/los términos de referencia deberían ser explícitos al respecto, como se ilustra en el ejemplo que figura en el *ANEXO A*.

El siguiente es un ejemplo de nota conceptual/términos de referencia que es **probable** que conduzca a una evaluación periódica satisfactoria:

- *“El objetivo de la misión de evaluación es redactar una amplio “Informe de desempeño en materia de GFP” preparado con arreglo a la metodología del PEFA a fin de ofrecer un análisis del funcionamiento general de los sistemas de GFP del país, así como un seguimiento de los progresos respecto de los indicadores del PEFA de la evaluación anterior que permitirán medir los cambios en el desempeño a través del tiempo.*
- *Los objetivos específicos de esta misión son:*
 1. *Actualizar el panorama general del desempeño en materia de GFP con arreglo al Marco para la Medición del Desempeño del PEFA.*
 2. *Establecer y explicar el nivel de mejora del desempeño sobre la base de las calificaciones de los indicadores del PEFA en comparación con los resultados de la evaluación anterior.*
 3. *Evaluar los resultados del estudio del PEFA, es decir, los cambios en el desempeño en relación con las actividades del proyecto y los posibles efectos en las calificaciones obtenidas.*
- *Se proporcionará a los evaluadores lo siguiente:*
 - a) *el informe final de la evaluación anterior;*
 - b) *todos los comentarios de los expertos evaluadores sobre la evaluación anterior;*
 - c) *si fuese necesario, el acceso al Jefe del equipo de la evaluación anterior para que explique los detalles de las calificaciones que se otorgaron.*

El siguiente es un extracto de nota conceptual/términos de referencia que es **poco probable** que conduzca a una evaluación reiterada satisfactoria:

- “*Existe la necesidad de hacer un balance de los progresos alcanzados desde la evaluación anterior*” (pero no se espera nada más específico de los evaluadores, como considerar las calificaciones anteriores).

Ejemplo: Nota conceptual/términos de referencia

Los objetivos específicos eran:

1. Actualizar el panorama general del desempeño en materia de GFP con arreglo al Marco para la Medición del Desempeño del PEFA.
2. Establecer y explicar el nivel de mejora del desempeño sobre la base de las calificaciones de los indicadores del PEFA en comparación con los resultados de la evaluación anterior.
3. Evaluar los resultados del estudio de PEFA, es decir, los cambios en el desempeño en relación con las actividades del programa de reformas y los posibles efectos en las calificaciones obtenidas.
4. Se debería tomar conocimiento de los motivos que pueden haber contribuido al cambio de calificaciones, tales como los siguientes:
 - cambios en las definiciones;
 - aumento de la disponibilidad de información o del acceso a ella;
 - diferencias en el muestreo y la agregación de la información;
 - distinto planteo en relación con los criterios profesionales;
 - inexactitudes en la metodología de calificación de la evaluación anterior, como la utilización de la calificación D cuando no se disponía de información suficiente para otorgar una calificación.
5. Los consultores deben asegurarse de que en el informe:
 - se expliquen todos los factores que repercuten en los cambios de calificación, indicador por indicador;
 - se identifique el cambio en el desempeño;
 - cualquier lector pueda seguir los cambios que se han producido desde la evaluación anterior.