



Marco de medición del desempeño en materia de gestión de las finanzas públicas

Informe de seguimiento 2010

Análisis de evaluaciones reiteradas, incluidos los cambios en el desempeño de sistemas de gestión de las finanzas públicas medidos mediante indicadores de gasto público y rendición de cuentas financieras

Secretaría del PEFA

Informe final

19 de mayo de 2011

Índice

Abreviaturas y siglas	4
Resumen	5
Capítulo 1: Introducción y metodología.....	11
Capítulo 2: Frecuencia de las evaluaciones reiteradas y razones para efectuarlas.....	15
2.1 Reseña	15
2.2 Conclusiones.....	15
<u>Capítulo 3: Eficacia en la medición de cambios en el desempeño.....</u>	18
3.1 Reseña	19
3.2 Conclusiones.....	19
Capítulo 4: Tendencias en el desempeño en materia de gestión de las finanzas públicas.....	31
4.1 Reseña	31
4.2 Conclusiones.....	31
Capítulo 5: Recomendaciones	39
Anexo A: Evaluaciones comparativas del gasto público y rendición de cuentas financieras	41
Anexo B: Evaluaciones reiteradas del gasto público y rendición de cuentas financieras no consideradas “comparativas”	43
Anexo C: Combinaciones de calificaciones de dimensiones entre una evaluación anterior y una evaluación reiterada.....	45
Anexo D: Países con evaluación inicial completada antes del 30 de junio de 2006 (es decir hace más de cuatro años y medio) que aún no han completado una evaluación reiterada...	46
Anexo E: Porcentaje de dimensiones de indicadores que pueden compararse con confianza entre distintas evaluaciones comparativas, por evaluación, si se prescinde de los casos “sin calificar”	47
Anexo F: Porcentaje de calificaciones que pueden compararse con confianza entre distintas evaluaciones comparativas para cada dimensión del indicador si se prescinde de los casos “sin calificar”	48
Anexo G: Clasificación porcentual de evaluaciones comparativas comparables, sin calificación y no comparables.....	50
Anexo H: Clasificación porcentual de cambios en las evaluaciones comparativas en los distintos indicadores y dimensiones	51
Anexo I: Metodología de codificación de la gestión de finanzas públicas elaborada por M. Andrews.....	53
Anexo J: Referencias.....	55

Los miembros del equipo básico encargado del informe fueron Helena Grandão Ramos (coordinadora de tarea), Clay Wescott y Brandon Lundberg. Proporcionaron informes técnicos Frans Ronsholt, Phil Sinnett, Charles Seibert y Tony Bennett. La revisión de pares del informe fue efectuada por Paulo de Renzio (participante independiente), Rachel Perrin (DFID), Mónica Rubiolo (SECO) y Steve Knack (Banco Mundial). El estudio fue supervisado por Frans Ronsholt.

Abreviaturas y siglas

APG	Apoyo presupuestario general
D-1, D-2, ...	Indicadores relativos a las prácticas de los donantes
DFID	Departamento de Desarrollo Internacional del Reino Unido
EA	Evaluación anterior
EC	Evaluación comparativa
ER	Evaluación reiterada
FMI	Fondo Monetario Internacional
GFP	Gestión de las finanzas públicas
HLG-1	Previsibilidad de transferencias de un más alto nivel de gobierno
ID	Indicador de desempeño
NORAD	Organismo Noruego de Cooperación para el Desarrollo
OADI	Organismo Australiano de Desarrollo Internacional
PEFA	Gasto público y rendición de cuentas financieras
PIB	Producto interno bruto
SECO	Secretaría de Estado para Asuntos Económicos de Suiza

Resumen

1. Una de las principales razones que llevaron a elaborar el marco del gasto público y rendición de cuentas financieras (PEFA, por su sigla en inglés) fue crear un instrumento de seguimiento que permitiera la medición de cambios en el desempeño de los sistemas de gestión de las finanzas públicas (GFP) a lo largo del tiempo a nivel de países. En el presente informe de seguimiento de la aplicación del marco del PEFA —el cuarto de ellos— se examina si se ha cumplido ese objetivo y se investigan tendencias de cambios globales de los sistemas de GFP sobre la base de un lote inicial de evaluaciones reiteradas del PEFA.
2. El informe abarca los informes de evaluación a los que se había dado forma final (informes finales) o que estaban sustancialmente completados (borradores completos de informes) al 6 de octubre de 2010 y que constituían evaluaciones de segunda o tercera generación del PEFA en un país (denominadas evaluaciones reiteradas [ER]). Se han realizado 45 ER en 38 países.
3. Treinta y tres de esas evaluaciones (de 29 países) reciben el nombre de “evaluaciones comparativas” (EC) porque están destinadas a medir cambios a lo largo del tiempo. Doce ER no tienen ese carácter porque su finalidad no es medir cambios ocurridos a partir de una evaluación anterior.

Frecuencia de las evaluaciones reiteradas y razones para efectuarlas

4. En los cinco años y cuatro meses transcurridos desde la puesta en acción del marco del PEFA se llevaron a cabo 206 evaluaciones del PEFA en 119 países. En alrededor de un tercio de ellos también se realizó una ER del PEFA. En un 25% de los 119 países esas ER fueron comparativas.
5. En promedio se realizaron EC 33 meses después de la evaluación anterior (EA), en un rango entre 1 y 4 años, lo que comprende siete países que realizaron EC transcurridos menos de 30 meses a partir de la EA, es decir un período más breve que el recomendado, de un intervalo mínimo de tres años. El promedio para esas siete EC fue de 22 meses.
6. Para los 26 países que realizaron EC no menos 30 meses después de la EA, las razones más importantes expresadas fueron medir el avance y contribuir al diseño o al seguimiento de un plan de acción o programa de reforma de la GFP; a continuación se mencionó el objetivo de facilitar el diálogo con los donantes y de establecer vínculos con apoyo presupuestario general (APG) actual o futuro. Cuando los intervalos fueron menores, las razones fueron fundamentalmente las mismas, pero se hizo más hincapié en el condicionamiento de la futura asistencia de los donantes a la realización de las EC, por lo cual es imperioso que se realicen al cabo de un breve intervalo.
7. Las 12 ER no comparativas se realizaron en promedio 30 meses después de la EA. Su principal finalidad consistió en preparar un parámetro de referencia, lo que lleva a pensar que los principales interesados no confiaban suficientemente en las evaluaciones anteriores.

8. No fue posible detectar ninguna modalidad de variación de la finalidad o la frecuencia de las EC basada en las características de los países o en el donante patrocinador.

9. Cinco países aún no han realizado EC transcurridos cuatro años y medio, o más, de la EA, pero en dos casos se está realizando una ER, o se prevé realizarla. Entre las razones que explican que los intervalos hayan sido largos cabe mencionar cambios de Gobierno y demoras en la puesta en marcha de un programa de reforma gubernamental, así como un arduo proceso de la evaluación inicial con escasa o ninguna adhesión del Gobierno.

10. Se recomienda que las ER del PEFA se realicen transcurridos entre tres y cinco años de la evaluación inicial. En conjunto, alrededor del 80% de las EC se realizaron con un intervalo de más de dos años y medio. Solo en cinco países no se han efectuado ER transcurridos cuatro años y medio de la primera evaluación, lo que lleva a pensar que ese tipo de evaluaciones se realiza en forma muy uniforme entre los distintos países y en gran medida dentro del intervalo estipulado.

Eficacia en la medición de cambios en el desempeño

11. Para poder utilizar las calificaciones de los indicadores del PEFA para medir los cambios a lo largo del tiempo es importante que pueda confiarse razonablemente en que cada par de calificaciones de la EA y la EC correspondientes a cada indicador (y dimensión del indicador) represente el cambio “real” en el desempeño. Se examinó cada una de las dimensiones de los indicadores correspondientes a cada una de las 33 EC para evaluar si la Secretaría del PEFA consideró adecuada la prueba para las calificaciones de la EA y la EC y para identificar algún otro problema de calificación (tal como el suministro de nuevas pruebas para la calificación de la EA, utilización de diferentes definiciones, o de diferentes mecanismos de toma de muestras, o diferentes interpretaciones de datos similares).

12. La indicación de “sin calificar” en la EA y/o en la EC es otro factor que impide medir el cambio a lo largo del tiempo. Pero como la ausencia de calificación en la dimensión de un indicador es fácilmente detectable al comparar pares de calificaciones, los usuarios de los datos de calificaciones sabrán dónde y cuándo reviste importancia ese factor.

13. Un nivel de comparabilidad no inferior al 80% para todas las dimensiones de un indicador —entre las calificaciones de la EA y la EC— se consideró satisfactorio cuando del conjunto de datos se excluyen las dimensiones “sin calificar”. Ese sólido nivel de comparabilidad fue alcanzado para 25 de las 33 EC, es decir para el 76% de los informes. En tres de las 33 EC (9%) se concluyó que el nivel de comparabilidad era insatisfactorio porque menos de la mitad de las calificaciones de las dimensiones de los indicadores pudo compararse con confianza.

14. Un análisis de los factores del proceso que pueden haber contribuido a la solidez de la medición de los cambios a lo largo del tiempo llevó a la conclusión de que todos los factores siguientes revestían importancia para lograr un alto nivel de comparabilidad entre la EA y la EC:

- Claridad de las notas conceptuales y los términos de referencia con respecto a la medición de los cambios en el desempeño desde la EA del PEFA.
- Utilización del mismo equipo de evaluación en ambas evaluaciones (EA y EC).
- Establecimiento de una estructura específica para la participación del Gobierno en la evaluación.
- Realización de talleres iniciales y finales para los principales interesados.
- Presentación de borradores de informes a la Secretaría del PEFA para recibir comentarios.

15. De los factores que anteceden, solo el examen de la Secretaría del PEFA de los borradores de los informes no estaba correlacionado con el grado de comparabilidad, por la sencilla razón de que todas las EC del análisis habían sido examinadas por la secretaría. No obstante, anteriores informes de seguimiento han demostrado que los exámenes a cargo de la secretaría contribuyen significativamente a mejorar el coeficiente de cumplimiento y, por lo tanto, la medida en que las calificaciones de los indicadores cuentan con adecuado respaldo probatorio. En consecuencia, en la solidez de la comparación de las calificaciones de las EA y las EC influyeron los exámenes de la secretaría.

16. Un análisis de la comparabilidad de cada una de las 74 dimensiones de los indicadores de todas las EC muestra que la comparabilidad media es del 82% de las observaciones si se excluyen los casos “sin calificar” (70% si se incluyen los casos “sin calificar”). El nivel de comparabilidad supera el 80% para 51 de las 74 dimensiones del marco. El más bajo nivel de comparabilidad (61%) corresponde al ID-7 i) “*magnitud del gasto no declarado del Gobierno*”, un indicador que es notoriamente difícil respaldar mediante pruebas sólidas. Se obtienen menores niveles de comparabilidad y diferentes rangos de dimensiones cuando en los conjuntos de datos se incluyen casos “sin calificar”. Este análisis puede contribuir a identificar dimensiones de indicadores que requieran aclaraciones y orientación adicionales para los evaluadores, inclusive haciendo hincapié en la realización de cursos de capacitación.

Tendencias del desempeño en materia de GFP

17. De todas las calificaciones de las dimensiones de los indicadores, un 11% de los casos mantuvieron calificaciones “A” que no pudieron mejorarse, en el 21% de los casos la calificación mejoró, en el 10% se mantuvieron las calificaciones “D”, que no pueden reducirse, en un 9% de los casos las calificaciones empeoraron, en un 21% de los casos se mantuvieron calificaciones “B” o “C”, y el 30%¹ de los casos no se prestaban a mediciones válidas del cambio. Se analizaron modalidades de cambio para *todos los indicadores o dimensiones en que era posible comparar válidamente el desempeño* desde una evaluación hasta la siguiente. Se cree que esos cambios representan principalmente variaciones reales del desempeño, aunque es imposible descartar que en algunos casos nuevas calificaciones puedan basarse en una mejor información o en diferentes juicios o interpretaciones, que no han podido detectarse en el actual proceso.

18. Para analizar las modalidades de desempeño se utilizó una metodología que categoriza los indicadores o dimensiones como atributos *formales* de la GFP cuando puede lograrse progreso a través de la adopción de leyes, reglamentos o instrumentos técnicos nuevos o mediante la focalización de tan solo unos pocos organismos, o en una

¹ El total supera el 100% debido al redondeo de las cifras. El 30% incluye los casos “sin calificar”.

etapa temprana del ciclo presupuestario, y atributos *funcionales* de la GFP cuando el progreso requiere la efectiva aplicación de nuevas leyes o reglamentos, la coordinación de numerosos organismos u ocuparse de etapas posteriores del ciclo presupuestario, es decir la clasificación de dimensiones de indicadores de acuerdo con las características conexas de la GFP, para distinguir entre situaciones *de jure* y *de facto*, etapas tempranas y posteriores y concentración/desconcentración de actores.

19. A partir del análisis se concluyó que la existencia de atributos formales genera una mayor probabilidad de mejoramiento o mantenimiento de la calificación máxima, en tanto que los atributos funcionales aumentan la probabilidad de empeoramiento o mantenimiento de la calificación máxima. En ambos tipos de atributos es mayor la proporción de calificaciones máximas y en aumento que de calificaciones mínimas y en proceso de empeoramiento, aunque las diferencias entre los atributos formales son mayores que entre los funcionales. Más específicamente:

- Los indicadores que representan concentración de actores muestran mejoras de desempeño mucho más pronunciadas que los que representan desconcentración de actores. En este último caso rara vez se pudo identificar alguna mejora global.
- Los indicadores que representan elementos iniciales y posteriores del ciclo presupuestario muestran resultados casi igualmente satisfactorios.
- Los indicadores de elementos *de jure* muestran grados de mejoramiento moderadamente mayores que los elementos *de facto*.
- Las diferencias entre calificaciones formales y funcionales son más pronunciadas en el caso de las calificaciones mínimas o en proceso de deterioro que en el caso de las máximas o en proceso de mejoramiento.

20. Se analizaron siete EC con intervalos de menos de 30 meses a partir de sus respectivas EA para buscar atributos distintivos en cuanto a medición del desempeño. Aunque se preveía que los intervalos de menos de tres años serían demasiado breves como para revelar el progreso, en realidad la proporción de calificaciones en aumento fue mayor que el promedio en los casos de intervalos breves, y la proporción de calificaciones no susceptibles de comparación fue más reducida. Ello puede explicarse quizá por el hecho de que en todos los casos de EC de intervalos breves un factor clave, entre los que motivaron la realización de la EC, fue que se trataba de una condición para obtener respaldo de donantes, lo que constituye un posible incentivo para mostrar el sistema de GFP del mejor modo posible. También pudo haber sido un incentivo para que el Gobierno y los donantes realizaran una temprana ER, si los interesados estaban convencidos de antemano de que una ER pondría de manifiesto un significativo mejoramiento de los sistemas de GFP.

Recomendaciones

21. En virtud de estas conclusiones se formularon recomendaciones, varias de las cuales se presentaron en anteriores informes de seguimiento. **La Secretaría del PEFA** debería adoptar las siguientes medidas:

- Revisar los documentos existentes relativos al proceso de evaluación (“Lista de comprobación de términos de referencia”, “Buenas prácticas en la aplicación del marco de referencia para la medición del desempeño en materia de GFP”, “Nota de orientación para las ER”) para destacar la necesidad de i) incluir en las notas conceptuales y los términos de referencia una referencia específica a las notas de orientación de la secretaría; ii) planificar la obtención de tiempo y recursos

- adicionales para analizar cambios; iii) presentar las notas conceptuales y los términos de referencia a la secretaría para que formule comentarios, y iv) lograr la transferencia de información detallada de la EA a los evaluadores de la EC, mediante la repetición de personas entre los evaluadores o la colaboración del jefe del equipo de la EA, y proporcionando a los evaluadores de la EC los comentarios de los interesados y de la Secretaría del PEFA sobre la EA.
- Revisar el material de capacitación existente para destacar los temas planteados en el punto anterior.
 - Examinar las dimensiones de los indicadores en que la comparabilidad es escasa para establecer si las dificultades en esa materia requieren aclarar el marco y la orientación para los evaluadores y determinar si debería considerarse la posibilidad de modificar los requisitos mínimos para la calificación de las dimensiones.

22. Los *organismos principales* deberían disponer lo necesario para aplicar ciertas buenas prácticas:

- En las notas conceptuales y los términos de referencia debería establecerse lo más específicamente posible lo que se espera de la ER, tal como se recomienda en la “Nota de orientación para las evaluaciones reiteradas” de la Secretaría del PEFA.
- Las notas conceptuales y los términos de referencia deberían especificar en forma suficientemente detallada la manera en que los evaluadores deberían incluir comparaciones con la EA en el informe de la EC y dar tiempo para verificar las razones que hayan llevado a adjudicar las calificaciones anteriores.
- En las notas conceptuales y los términos de referencia debería disponerse que el equipo de evaluación registre toda la evaluación pertinente (por ejemplo en un CD) en forma comprensible y de fácil acceso para otros expertos en una ER ulterior.
- Las notas conceptuales y los términos de referencia deberían incluirse como anexo del borrador o del texto final de los informes de evaluación.
- Las notas conceptuales y los términos de referencia deberían ser objeto de un proceso de garantía de calidad (revisión de pares), a cargo de la Secretaría del PEFA, como garantía de la adecuada implementación de las recomendaciones que anteceden.
- Debería seguirse la práctica de proporcionar al equipo de evaluadores de la EC toda la información y los documentos pertinentes surgidos de las evaluaciones anteriores, incluidos el informe final y los comentarios de los interesados y de la secretaría.
- Los organismos principales deberían facilitar el contacto entre los jefes de equipos de EA es y actuales, aunque ello requiera más tiempo e implique costos adicionales.
- Deberían proporcionarse a la secretaría razones específicas por las que no se compartan sus comentarios a las notas conceptuales y los términos de referencia, así como las evaluaciones.

23. Al realizar una ER, *los evaluadores* deberían tener en cuenta las siguientes recomendaciones:

- Seguir el asesoramiento contenido en la “Nota de orientación para las ER” de la Secretaría del PEFA.
- Solicitar al administrador de la evaluación (organismo principal) información sobre la EA (borradores de informes e informes finales, y comentarios del proceso de examen de calidad).
- Solicitar al organismo principal que se ponga en contacto con el equipo de EA.

Labor adicional

24. Las conclusiones y recomendaciones que anteceden constituyen un “aperitivo” de lo que puede aportar la ampliación de la base de datos de ER del PEFA para profundizar el conocimiento de los cambios en el desempeño de la GFP. A lo largo del informe se formulan numerosas sugerencias sobre la manera en que una *investigación más profunda* puede fortalecer las conclusiones o ahondar el conocimiento de cuestiones básicas. Si bien la Secretaría del PEFA tiene la intención de seguir realizando en parte esa labor, especialmente en cuanto guarda relación con la necesidad de desarrollo y orientación sobre el marco del PEFA y su aplicación, muchas otras oportunidades están a disposición de los investigadores, en general en cuanto a las tendencias en el desempeño de los sistemas de GFP.

Capítulo 1: Introducción y metodología

25. Este es el cuarto informe de seguimiento preparado por la secretaría. La medición del progreso logrado a lo largo del tiempo es uno de los objetivos principales del marco del PEFA. Hasta ahora la mayoría de las evaluaciones realizadas han sido evaluaciones iniciales, pero el número de ER es cada vez mayor, ya que la cobertura de la base de referencia está casi completa en algunas regiones, y las evaluaciones iniciales tienen ya no menos de tres años.

26. Uno de los objetivos de una ER consiste en medir el desempeño desde la EA. En una ER se examinan los cambios específicos experimentados en el desempeño de un sistema verificando lo que ha cambiado y en qué medida. Cuando se utiliza sistemáticamente, los interesados pueden esperar que la aplicación reiterada del marco proporcione pruebas sobre la magnitud del cambio experimentado por el desempeño en la GFP de un país. En el informe de desempeño en materia de gestión de las finanzas públicas se reconocerán además los esfuerzos realizados por el Gobierno para reformar su sistema de GFP mediante la descripción de reformas recientes y en curso, aunque aún no hayan influido en el desempeño de la GFP. Se advierte sin dificultad que dos evaluaciones con cierta distancia en el tiempo podrán compararse si representan aplicaciones sistemáticas de la metodología. Si la evaluación inicial supone la aplicación rigurosa de la metodología lo ideal es que los interesados sepan que el avance se ha medido con precisión a lo largo del tiempo.

27. Basándose en el considerable número de ER actualmente disponibles, la finalidad principal del *Informe de seguimiento 2010 (IS 10)* consiste en establecer si el marco del PEFA permite medir de manera confiable los cambios en el desempeño a lo largo del tiempo². Los resultados del análisis revestirán gran importancia para establecer si se requiere una orientación técnica y de procesos adicional. El análisis permitirá asimismo dar orientación sobre comparación de calificaciones de indicadores del PEFA de diferentes años, así como proporcionar un insumo importante para la evaluación externa global del programa PEFA.

28. En el informe se procura dar respuesta a tres preguntas principales: i) ¿con qué frecuencia y por qué motivos se realizan ER?; ii) ¿el marco es un instrumento eficaz para medir los cambios, y estos podrían medirse con mayor validez y confiabilidad?, y iii) ¿qué tendencias en materia de desempeño de la GFP revelan las ER?

29. A los efectos del presente informe el equipo consideró que debe entenderse:

i) Por “*evaluación anterior*”, la que mida los sistemas y el desempeño del proceso de la GFP del país en cierto momento, siguiendo la metodología del PEFA. El objeto de la evaluación puede consistir en establecer una base de referencia frente a la cual pueda medirse el progreso en el desempeño en materia de GFP a través de indicadores o dimensiones³ a lo largo de años futuros.

² El estudio no incluyó un análisis de la totalidad del informe de evaluación, por no ser ese su objetivo, sino que centró la atención en temas específicos de las secciones 1 y 2 de la evaluación resumida, y en la calidad del atributo de comparabilidad de indicadores y dimensiones que aparece en la sección 3.

³ En el presente informe se utilizan calificaciones de dimensiones, salvo cuando un indicador no tiene diferentes dimensiones, en cuyo caso se utiliza la calificación del indicador.

ii) Por “*evaluación comparativa*” (EC) una ER del PEFA en que se mencione la EA, se comparen las calificaciones de las dos evaluaciones y se intente explicar las diferencias y cambios en el desempeño de la GFP.

30. Basándose en esas definiciones, en el informe se consideraron 33 ER (anexo A), a las que se da el nombre de “*evaluaciones comparativas*” porque a través de ellas se procura medir los cambios a lo largo del tiempo. Se descalificaron 12 ER porque no cumplían los criterios de una EC (véase el anexo B, donde aparece una breve explicación para cada caso).

31. En el capítulo 2 se analiza la frecuencia de las ER, y las razones que las motivan. En el capítulo 3 se evalúa la medida en que el marco permite medir eficazmente los cambios. En el capítulo 4 se resumen las tendencias en materia de desempeño de la GFP, reveladas por ER, y el capítulo 5 contiene recomendaciones.

32. La siguiente es la metodología utilizada para los capítulos 2, 3 y 4:

33. En el **capítulo 2** se consideran las razones que llevan a realizar una EC, tal como se enuncian en los informes, y se pregunta a los interesados que hayan financiado o llevado a cabo EC del PEFA si hubo razones adicionales. Se tienen en cuenta los borradores, los informes finales y los exámenes y comentarios de seguimiento de la secretaría con respecto a las secciones 1 y 2 de la evaluación resumida. Se mantuvieron comunicaciones telefónicas y por correo electrónico con numerosos interesados que han financiado o llevado a cabo ER del PEFA (representantes de organismos principales y jefes de equipos de trabajo). Se obtuvieron respuestas en alrededor del 70% de los casos. En algunos casos los interesados entrevistados expresaron opiniones diferentes a las contenidas en los informes escritos, por lo cual fue necesario interpretar racionalmente la situación actual. Por ejemplo, en muchos casos es difícil establecer si los donantes o el país en cuestión habían formulado la solicitud inicial de realización de una EC, y cuáles eran las razones principales a las que obedecía la solicitud. En el análisis se consideraron las 33 EC a las que se había dado forma final o que estaban sustancialmente completadas. En los casos en que los países han realizado más de una EC el análisis se centró en la más reciente, salvo en el caso de Malawi, en que se analizaron dos EC.

34. En relación con los capítulos 3 y 4, la base principal del análisis del progreso en una EC es una evaluación de indicadores o dimensiones calificados. Hay 16 posibles observaciones comparativas entre las dimensiones calificadas, que se presentan en el anexo C. Si bien el desempeño, en las diferentes dimensiones, puede mejorar, mantenerse o deteriorarse, dadas las cuatro posibilidades de calificación (“A”, “B”, “C” o “D”), las combinaciones de observaciones comparativas permiten un examen más profundo de la variación del desempeño entre una EA y la que la siga.

35. En algunos análisis incluidos en el presente informe se utilizan también dimensiones que no se califican (conocidas como “sin calificar”), pero como no es posible realizar una comparación entre dos evaluaciones cuando alguna de ellas no ha sido calificada, se consideran separadamente⁴.

⁴ Muchas razones pueden explicar que no se haya calificado a un indicador o dimensión. La Secretaría del PEFA elaboró una metodología de abstención de la calificación, basada en la observación de tres condiciones principales: “sin calificar”, cuando no se dispone de información suficiente para calificar un

36. En la mayoría de las partes del informe las observaciones comparativas se expresan como porcentaje del número total de observaciones, incluidas las que no están calificadas, pues se prevé que el número de observaciones no calificadas se vaya reduciendo a medida que mejoren los sistemas de GFP y las técnicas de evaluación del PEFA. Esa es la modalidad observada en la muestra, en que el porcentaje de observaciones no calificadas y las que no poseen comparabilidad por otras razones pasó de un promedio del 29% de todas las EA al 20% para las EC.

37. En el **capítulo 3** se examinan las EC para establecer si los indicadores y dimensiones pueden compararse con confianza con los de la EA. Como ya se señaló, se utilizan calificaciones de dimensiones, salvo en los casos en que un indicador no tiene dimensiones, y en ese supuesto se utiliza la calificación del indicador. Esto implica analizar las pruebas proporcionadas para cada indicador y dimensión evaluados para tener la certeza de que en términos generales miden lo mismo. Se tuvieron en cuenta los indicadores de informes de evaluaciones anteriores y comparativas, los comentarios y exámenes de seguimiento de la secretaría sobre borradores, e informes finales correspondientes a la EA y a la comparativa. Si no existía un examen de seguimiento, este se realizaba. Se identificaron cinco “*problemas de seguimiento*” que impiden realizar comparaciones confiables: a) nuevas pruebas recogidas que no estaban disponibles en la EA, b) cambios de definiciones, c) diferentes interpretaciones, d) diferentes procedimientos de toma de muestras, y e) nuevas calificaciones de evaluaciones anteriores en la ER.

38. Se tuvieron en cuenta las notas conceptuales y los términos de referencia⁵ y la información disponible proveniente de los informes y de los documentos internos de la secretaría sobre factores de procesos que puedan afectar a la calidad de la medición de los cambios, a saber, el uso del mismo equipo de evaluación para las evaluaciones, la participación del Gobierno en la ER, talleres iniciales y de salida y presentación de borradores de informes a la Secretaría del PEFA para obtener comentarios. También se tuvieron en cuenta las notas conceptuales y los términos de referencia que han sido preparados y la ER correspondiente desde que la Secretaría del PEFA publicó la nota de orientación “*Para las evaluaciones periódicas*”, en febrero de 2010.

39. Sobre la base de ese análisis, en el capítulo tres se evalúa en qué medida el marco facilita la medición de cambios en algunos aspectos y las razones por las que en otros aspectos no la facilita. En el análisis se consideraron 33 EC, 22 de las cuales se realizaron en su totalidad (informe final) y 11 estaban sustancialmente completadas (borrador completo del informe).

40. En el capítulo 4 se examina, en los casos en que calificaciones de indicadores y dimensiones pueden compararse con confianza con evaluaciones anteriores, qué está cambiando y en qué medida, y qué indicadores están cambiando. Además se examinaron las modalidades de cambios entre diferentes tipos de indicadores y características de países y en diferentes regiones. Para el análisis se utilizó una metodología (Andrews,

indicador; “no aplicable”, cuando un indicador o una dimensión no se aplica al país que se evalúa, y “no utilizado”, cuando no se tiene intención de calificar al indicador o a la dimensión.

⁵ Menos de la mitad de las notas conceptuales y los términos de referencia finales estuvieron a disposición de la secretaría al comienzo de la preparación del *IS 10*. Las notas conceptuales y los términos de referencia faltantes se obtuvieron directamente de los interesados, de informes finales y en un caso del consultor. Dos notas conceptuales y términos de referencia no se prepararon.

2009; Porter y otros, 2010) en que se clasifica a los indicadores o dimensiones en tres parejas: *de jure* y *de facto*, iniciales y posteriores, y concentrados y desconcentrados⁶.

- En la primera pareja se contraponen las dimensiones del PEFA en que una calificación no inferior a “C” podría obtenerse a través de la sanción de una nueva ley o el anuncio de una nueva práctica, aunque no se implementen (es decir una reforma *de jure*) con dimensiones cuyas calificaciones requieran la implementación real, o un compromiso significativo (*de facto*).
- En la segunda pareja se contraponen dimensiones del PEFA tales como presupuestación estratégica (previsión plurianual, planificación estratégica, planificación de inversiones, planificación de la deuda), preparación del presupuesto anual, análisis parlamentario del presupuesto anual y la estructura de los documentos presupuestarios formales por una parte (**iniciales**), y gestión de recursos (inclusive gestión de entradas y salidas de efectivo, adquisiciones, nómina); control interno, auditoría y seguimiento internos, contabilidad e información, auditoría externa y análisis parlamentario de informes de auditoría, por otra parte (**posteriores**).
- En la tercera pareja se contraponen los aspectos del PEFA que están bajo el control de órganos centrales, reguladores, como el Ministerio de Finanzas (**control concentrado**), con aquellos en que es preciso comprometer a una pluralidad de órganos o autoridades subnacionales (**control desconcentrado**).

41. El cambio a lo largo del tiempo se mide principalmente por su orientación, y no por el número de pasos ascendentes o descendentes en la escala de calificaciones. Los cambios se agregan luego en las diferentes dimensiones de los indicadores y en los diferentes países, dando a todas las dimensiones igual ponderación y haciendo lo propio con los países. Para la preparación de este informe no se realizaron análisis de sensibilidad para poner a prueba los resultados de modificaciones de este enfoque. Será útil realizar una labor adicional a ese respecto para confirmar la solidez de las conclusiones.

⁶ Esta metodología se ha utilizado para algunos proyectos de estudio ampliamente difundidos. Si bien la Secretaría del PEFA no está totalmente de acuerdo con la utilización de esa clasificación para todos los indicadores, la metodología se utiliza aquí sin ninguna modificación ya que un desarrollo adicional de este instrumento llevaría tiempo y en él deberían participar los creadores originales de la metodología. Podría considerarse la posibilidad de una labor adicional encaminada a mejorar las clasificaciones de los indicadores e investigar el impacto que puedan tener sobre los resultados del análisis.

Capítulo 2: Frecuencia de las evaluaciones reiteradas y razones para efectuarlas

2.1 Reseña

42. En esta sección se examinan las siguientes cuestiones:

- ¿Cuál fue la frecuencia de las ER?
- ¿Qué razones llevan a realizar EC, cuándo se realizan con gran frecuencia y cuándo se hacen en cumplimiento del intervalo recomendado de tres a cinco años?
- ¿Por qué algunos países no han realizado ER dentro del intervalo recomendado?
- ¿Qué otras finalidades, además de la medición del cambio en el desempeño, han llevado a realizar EC?
- ¿Las finalidades varían según las características de los países o de los propósitos del donante patrocinador?
- ¿Alguna característica del país contribuye a que las EC se realicen oportunamente?

2.2 Conclusiones

43. Entre junio de 2005 y el 6 de octubre de 2010 se llevaron a cabo 45 evaluaciones mediante la utilización del marco tanto para la primera como para las subsiguientes evaluaciones. Ellas abarcan 41 Gobiernos nacionales y subnacionales de 38 países (tres ER subnacionales se llevaron a cabo en Etiopía, además de una ER nacional). De esas 45 ER (cuya lista aparece en el anexo A), 33 cumplieron los criterios que les dan carácter de EC; no así las 12 restantes (cuya lista aparece en el anexo B).

44. Las EC se realizaron, en promedio, 33 meses después de la EA, en un intervalo de uno a cuatro años⁷. Siete países llevaron a cabo EC menos de 30 meses después de la EA, lo que hace un promedio de 22 meses⁸. Se trata de un intervalo más breve que el recomendado por la directriz (de tres a cinco años), establecido debido al tiempo que se presumía necesario para que las reformas pusieran de manifiesto resultados medibles. Un factor que debe tenerse en cuenta es que la distancia entre las fechas de la EA y de la EC puede no reflejar necesariamente la que existe entre los conjuntos de datos básicos utilizados. En el caso de Kenya (30 meses entre una misión y otra), por ejemplo, la EA utilizó datos correspondientes al ejercicio de 2004-05, en tanto que la EC utilizó datos correspondientes al de 2007-08; por lo tanto se cumplió el intervalo recomendado de tres años entre dos evaluaciones consecutivas.

45. Teniendo presente lo que antecede, las dos razones más comúnmente enunciadas para el PEFA en los casos de intervalos breves fueron la necesidad de facilitar el diálogo con los donantes y la de medir el progreso de la reforma. En Afganistán, por ejemplo, el Banco Mundial, que financia alrededor del 30% del presupuesto no militar del Gobierno,

⁷ Datos basados en las fechas de las principales misiones correspondientes a cada evaluación. En tres casos en que no se conocía la fecha de la misión principal de la EA se utilizó la fecha del informe.

⁸ Datos basados en las fechas de las principales misiones correspondientes a cada evaluación.

necesitaba una EC como insumo para su marco fiduciario. En la República Dominicana una de las finalidades fue lograr mayor uniformidad en el diálogo con la comunidad internacional; no obstante, también fue crucial la preocupación del propio Gobierno en cuanto al seguimiento del progreso en la gestión de los recursos. En Mozambique se consideró que la EC era una base sólida para evaluar el progreso a lo largo del tiempo. En Tanzania una de las finalidades expresadas fue la de "...sentar las bases para un diálogo entre el Gobierno y los donantes sobre futuras reformas de la GFP".

46. Otras razones expresadas fueron la necesidad de que el PEFA contribuya al diseño o seguimiento del plan de acción o programa de reforma de la GFP y sea un requisito para un respaldo presupuestario actual o futuro o un futuro proyecto de GFP. Por ejemplo, la Unión Europea persuadió al Gobierno de Malawi para que postergara hasta que se conocieran los resultados de la primera EC la definición de un plan de acción sobre reforma de la GFP, que fue entonces la base de un significativo incremento del APG. La más reciente EC de Malawi (la segunda de ese país) sentó las bases de un segundo plan de acción de GFP que fue, asimismo, una condición para recibir APG de la Unión Europea. El Gobierno dio creciente atención a las sucesivas evaluaciones del PEFA: la primera no fue considerada adecuadamente; la segunda se tomó más en serio debido a su vínculo con el APG, y la tercera se debatió en todos sus aspectos dentro del Gobierno y contribuyó a un diálogo útil entre este y los donantes de APG. Los interesados reconocieron el hecho de que el intervalo de un año entre la primera EA y la primera EC era demasiado breve como para que pudiera percibirse algún progreso; el Gobierno tenía expectativas mucho más ambiciosas que los donantes de que la primera EC mostraría mejoras significativas. Finalmente, en Kosovo significativos cambios del marco institucional y constitucional del país llevaron al Gobierno y a los donantes a reclamar una evaluación actualizada en que se tuvieran en cuenta esos cambios.

47. En relación con las 26 EC realizadas 30 meses⁹ o más (promedio: 37 meses) después de la EA, las más importantes razones enunciadas consistieron en medir el progreso y contribuir al diseño o al seguimiento de un plan de acción o programa de reforma de la GFP; a continuación se mencionó el objetivo de facilitar el diálogo con los donantes y el establecimiento de vínculos con un APG actual o futuro. Dominica, por ejemplo, realizó una EC porque el Gobierno consideró que la EA era débil y había sido realizada de manera insatisfactoria, y la EC le daría una base sólida para elaborar un nuevo plan de acción de GFP y además se utilizaría para determinar la continua elegibilidad para recibir respaldo presupuestario. En Uganda la EC contribuyó a un examen anual del desempeño de la GFP, a la reducción de los costos de transacciones públicas, lo que facilitó el diálogo con los donantes en torno a un acervo de información acordado; sentó bases decisivas para el APG y un mecanismo de confrontación con la realidad acerca de si se había avanzado en la GFP en respuesta al respaldo recibido de un cuantioso fondo común de GFP. En Timor Leste se llevó a cabo una EC como parte de una evaluación de la transparencia fiscal en relación con las disposiciones del código de transparencia fiscal del Fondo Monetario Internacional (FMI), un informe sobre la observancia de los códigos y normas en materia fiscal, para evaluar el desempeño y determinar la necesidad de reorientar los programas de GFP. En la República Kirguisa, a través de una EC se verificó el avance en las reformas de la GFP y se respaldó el diseño de un enfoque más estratégico y programático para las reformas de la GFP a través de una visión de mediano plazo de la GFP. En Tonga, el Ministerio de Finanzas tenía interés en conocer las opiniones independientes de personas sin conocimiento anterior del país

⁹ Datos basados en informes de las principales misiones.

sobre las reformas que serían razonables, manejables y provechosas, y las que no. En Ghana un nuevo Gobierno deseaba obtener un cuadro preciso de la situación de la GFP y alentar a los asociados de la ayuda para utilizar más plenamente los sistemas del país.

48. Además hubo 12 ejemplos de evaluaciones que siguieron a evaluaciones anteriores que no se consideran EC debido a que en ellas no se reconoció la EA ni analizaron los cambios en el desempeño de la GFP. Si bien la razón para ello guarda relación específica con cada evaluación, en el cuadro 1 se señala que los problemas de calidad pueden ser especialmente pronunciados para la EA o la ER. En esos 12 casos aumentó en promedio el número de dimensiones sin calificar en las EA y disminuyó la tasa de revisión por parte de la Secretaría del PEFA¹⁰. Esas ER se realizaron, en promedio, 30 meses¹¹ después de la anterior evaluación. La finalidad principal consistió en preparar una base de referencia, lo que lleva a pensar que la EA no inspiraba suficiente confianza. Por ejemplo, la EA correspondiente a Bolivia había sido una autoevaluación que abarcaba exclusivamente una parte del marco, por lo cual se necesitaba una evaluación completa para producir una base de referencia completa. En el caso de la República Centroafricana, el Banco Africano de Desarrollo estaba preparando un programa de respaldo institucional para la GFP y necesitaba un PEFA actualizado como referencia. Hubo objeciones, porque la EA no había sido validada en un taller. Además el Banco Africano de Desarrollo y el APG del Banco Mundial utilizaron indicadores del PEFA como indicadores de desempeño y necesitaban una evaluación actualizada para preparar el informe de culminación. En el caso de Montserrat la EA de 2008 fue considerada oportuna, pues se basaba en gran medida en la evaluación de riesgo fiduciario del Departamento de Desarrollo Internacional del Reino Unido (DFID) de 2006. Además, la EA de 2008 no había sido sometida al examen de la Secretaría del PEFA, lo que impidió al DFID utilizarla como base para su evaluación de riesgo fiduciario de 2010 (que requieren las directrices internas del DFID), por lo cual era necesaria una nueva EC.

Cuadro 1: Diferencias entre evaluaciones comparativas y otras evaluaciones reiteradas en materia de problemas de cobertura y garantía de calidad

Evaluaciones reiteradas	Promedio de dimensiones sin calificar		Borrador de la evaluación de la Secretaría del PEFA revisado (porcentaje)	
	EA	ER	EA	ER
33 EC incluidas	10	3	100%	100%
12 ER excluidas	27	3	83%	83%

49. Se realizó un análisis de correlaciones para establecer si el objeto o la frecuencia de las EC variaba en función de las características del país o del donante patrocinador, pero no fue posible identificar modalidades. Se tuvo en cuenta la necesidad de utilizar regresiones múltiples para tratar de encontrar resultados posibles. No obstante, dada la pequeñez de la muestra, lo limitado de la gama de la variable dependiente (intervalo entre evaluaciones), y el carácter subjetivo de las posibles variables explicativas (por ejemplo la presión que se ejercía para poner de manifiesto el progreso de las reformas de la GFP), se concluyó que no era probable que ese análisis arrojara resultados significativos en esta etapa.

¹⁰ En el anexo B aparecen las razones y la lista completa de ER excluidas.

¹¹ Según datos de cobertura de los informes.

50. En total 23 países habían completado una evaluación inicial al final de junio de 2006, es decir en el primer año que siguió a la puesta en marcha del marco. Casi cinco años después, 18 de esos países (el 78% de ellos) habían llevado a cabo una ER. Algunas de esas ER (seis) no eran EC por razones que más arriba se exponen. Se examinaron las cinco evaluaciones restantes (ref.: anexo D) para comprender las razones por las que no se había llevado a cabo una ER dentro del intervalo recomendado. La EA de Fiji finalizó en junio de 2005. La falta de una ER puede explicarse por razones vinculadas con la demanda —se afirma que el Gobierno no consideraba satisfactoria dicha EA— y con la oferta, ya que los donantes habían reducido la asistencia desde el advenimiento del régimen militar, en 2006. Otro vacío prolongado fue el de Bangladesh, ya que el Gobierno estaba poco identificado con el PEFA realizado en 2005, por lo cual se utilizó una metodología modificada para el PEFA, lo que a su vez hizo dudar de la validez del proceso y el Gobierno no aceptó como creíbles las calificaciones, y el informe no se publicó, con excepción de un cuadro resumido como anexo de la estrategia de asistencia al país del Banco Mundial. Desde entonces se han realizado diagnósticos de la GFP que cubren algunos ámbitos del PEFA (Gobierno de Bangladesh y DFID, 2007; Banco Mundial, 2006, 2007, 2009; DFID, 2009). Actualmente se está realizando una EC para Bangladesh. En el caso de Siria, en un informe de asistencia técnica del FMI se incluyó una evaluación parcial del PEFA. No existió un acuerdo explícito con el Gobierno sobre la utilización de la metodología del PEFA, con la que el Gobierno no estaba familiarizado. Evidentemente la falta de identificación del Gobierno con la evaluación representó también un problema para la evaluación en la República del Congo (Brazzaville). En el caso de Uganda (Gobierno local), se programó la realización de una ER en 2011 y la falta de identificación del Gobierno quizá no contribuyó a la demora ocurrida, ya que entretanto se habían realizado varias ER a nivel nacional. En conjunto, una importante razón que llevó a no realizar una ER dentro del período quinquenal recomendado parece haber sido la escasa identificación del Gobierno con la evaluación inicial del PEFA. Tal vez influyeron varios otros factores (como el grado de dependencia de la ayuda y la política local), pero el número de casos es demasiado reducido como para que se justifique un análisis más minucioso.

51. En conclusión, la gran mayoría de las evaluaciones del PEFA son seguidas por ER dentro del intervalo de tres a cinco años recomendado (o un intervalo levemente más breve). La mayoría de las 33 EC se realizaron no más de 30 meses después de la EA. Las más importantes razones expresadas para realizar las EC fueron el propósito de medir el progreso y contribuir al diseño o al seguimiento de un plan de acción o programa de reforma de la GFP, seguidas por la facilitación del diálogo con los donantes y el establecimiento de vínculos con APG en curso o futuro. Se realizaron siete EC antes de que transcurrieran 30 meses de la EA. En la mayoría de los casos esa decisión fue motivada por la necesidad inmediata de los donantes y/o de los Gobiernos de medir el progreso alcanzado. No existieron modalidades discernibles, en cuanto al objeto y a la frecuencia de las EC, vinculadas con características de los países o con los donantes patrocinadores. Los 12 países que realizaron ER —no consideradas EC— no reconocieron la EA ni analizaron los cambios en el desempeño de la GFP. Preocupaciones con respecto a la calidad de la EA pueden haber influido poderosamente a ese respecto. Finalmente, se registraron cinco casos en que no se habían realizado ER entre cinco y seis años después de las EA. Una importante razón parece ser la escasa identificación del Gobierno con el proceso del PEFA.

Capítulo 3: Eficacia en la medición de cambios en el desempeño

3.1 Reseña

52. En esta sección se examinan las siguientes cuestiones:
- ¿Con qué calidad se miden los cambios?
 - En relación con cada caso, ¿qué calificaciones de indicadores y dimensiones pueden compararse con confianza con calificaciones anteriores para determinar cambios, basados en pruebas?
 - ¿Cuáles son las razones de que las calificaciones de indicadores y dimensiones no puedan compararse con confianza con calificaciones anteriores (por ejemplo calidad de las pruebas, explicación y escala del cuestionamiento de calificaciones anteriores; claridad de los términos de referencia con respecto a la medición del progreso a lo largo del tiempo)?
 - ¿Cuáles son los “factores de procesos” que afectan a la calidad de la medición de los cambios?

3.2 Conclusiones

Contexto

53. Para analizar los cambios en el desempeño a lo largo del tiempo en el presente informe se consideraron exclusivamente 33 ER a las que se denomina “evaluaciones comparativas”. Al 6 de octubre de 2010 se había dado forma final a 22 EC (informe final), 11 de las cuales estaban sustancialmente completadas (borrador completo del informe). De las 22 EC finalizadas, 10 eran de acceso público¹². Las 33 EC abarcaban 29 países: 14 de África (AFR), seis de América Latina (LAC), cuatro de Asia oriental y el Pacífico (EAP), cuatro de Europa oriental y Asia central (ECA) y una de Asia meridional. No se ha realizado ninguna EC en países de Oriente Medio y Norte de África (MENA) aunque han transcurrido tres años desde que se realizó en algunos países la última evaluación del PEFA (por ejemplo en Siria).

54. La principal base del análisis del progreso es una evaluación de los indicadores o dimensiones que han sido calificados¹³. La comparabilidad se analiza sobre la base de un “análisis vertical” (en términos de comparabilidad de las calificaciones de una EC con su EA en todas las dimensiones del indicador) y de un “análisis horizontal” (en términos de comparabilidad de determinada dimensión del indicador en todas las EC).

55. La adjudicación de “sin calificar” a una evaluación, o a ambas, hace imposible comparar las calificaciones de dos evaluaciones. Como se señala en el capítulo 1, los indicadores o dimensiones que no reciben calificación también se utilizan en el análisis en el presente informe, pero como no es posible comparar la EA y la EC se consideran separadamente. Se trata de una cuestión importante para explicar la imposibilidad de

¹² Disponibles en www.pefa.org: Ghana, Guinea Bissau, Kenya, República Kirguisa, Moldova, Timor Leste, Trinidad y Tabago, Tonga y Uganda. En enero de 2011 se publicó una EC adicional, correspondiente a la República Dominicana.

¹³ Solo se analizaron calificaciones de dimensiones. Algunos indicadores tienen solo una dimensión, en cuyo caso las calificaciones de la dimensión y del indicador coinciden.

realizar comparaciones. En los cuadros 2, 3 y 3a se ilustra el impacto del factor “sin calificar” en el análisis vertical y en el horizontal. Cuando en el conjunto de datos se incluyen casos “sin calificar” el nivel de comparabilidad en las 74 dimensiones es del 70%. Cuando esos casos se excluyen, el nivel de comparabilidad se incrementa hasta llegar al 82%. Esto significa que el 30% de los casos de imposibilidad de realizar comparaciones se deben al factor “sin calificar”, lo que se señala claramente en los informes de evaluación. Una diferencia similar se produce en materia de comparabilidad entre EC individuales.

Cuadro 2. Comparabilidad entre evaluaciones comparativas

<i>Banda de comparabilidad de calificaciones</i>	<i>Banda 80% a 100%</i>	<i>Banda 50% a 79%</i>	<i>Banda ≤ 50%</i>	<i>Total de EC</i>
Número de EC que incluyen casos “sin calificar” comprendidos dentro de cada banda	15	11	7	33
Número de EC que excluyen casos “sin calificar” comprendidos dentro de cada banda	25	5	3	33

Cuadro 3. Comparación entre indicadores/dimensiones

Porcentaje de indicadores/dimensiones que pueden compararse con confianza cuando se incluyen casos “sin calificar”	70%
Porcentaje de indicadores/dimensiones que pueden compararse con confianza cuando se excluyen casos “sin calificar”	82%

Cuadro 3a. Comparación entre dimensiones de un mismo indicador (clasificación por banda de comparabilidad y número de indicadores)

<i>Banda de comparabilidad de calificaciones</i>	<i>Banda 80% a 100%</i>	<i>Banda 60% a 79%</i>	<i>Banda 50% a 59%</i>	<i>Banda ≤ 49%</i>	<i>Total de indicadores y dimensiones</i>
Número de indicadores cuando se incluyen los casos “sin calificar”*	19	43	7	5	74
Número de indicadores cuando se excluyen los casos “sin calificar”*	51	23	0	0	74

*Se eliminaron tres dimensiones del indicador previsibilidad de transferencias de un más alto nivel de gobierno (HLG-1). Véase la explicación en el párrafo 53.

*Comparabilidad entre evaluaciones comparativas (análisis vertical)*¹⁴

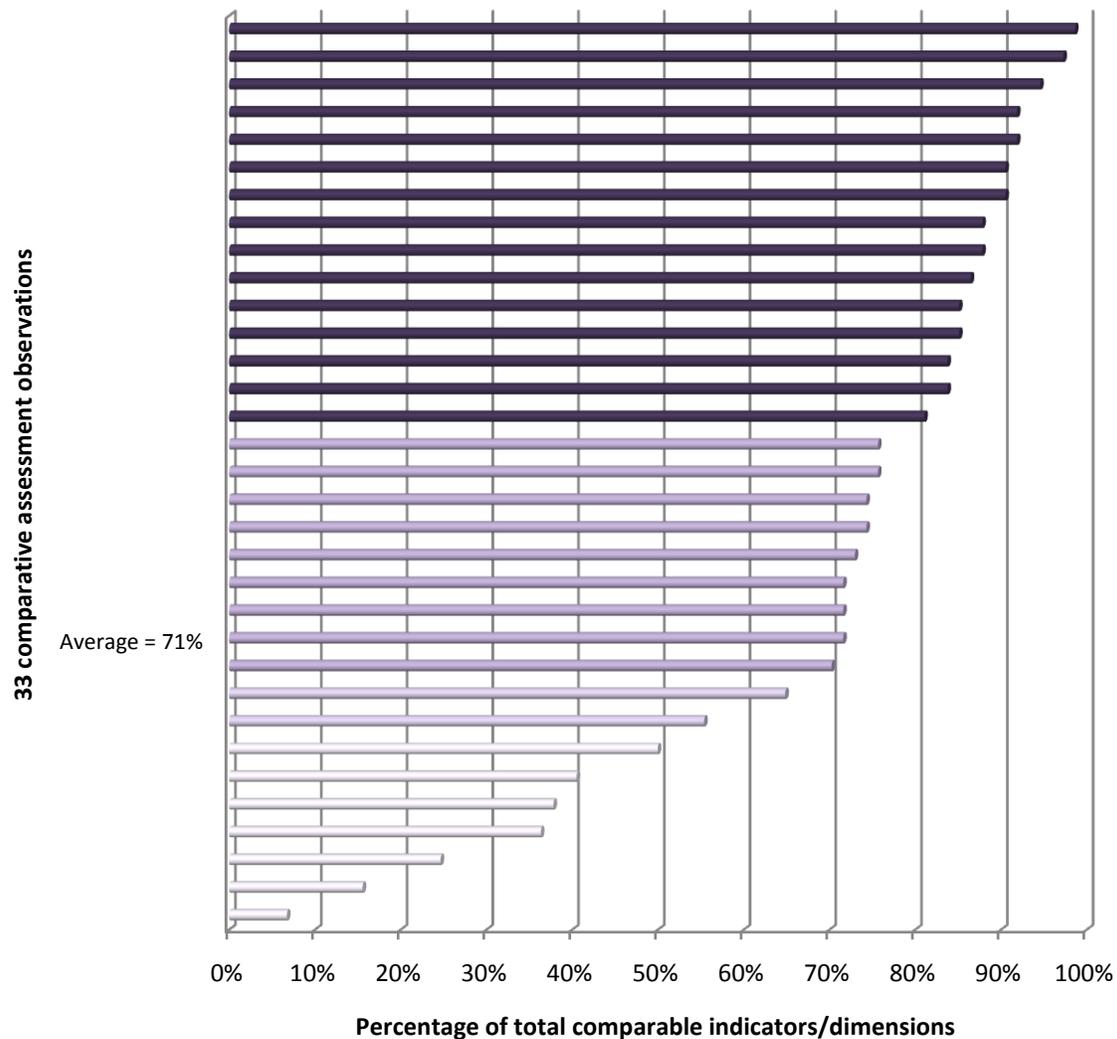
56. Al examinar las 33 EC, los datos muestran que en 15 de ellas entre el 80% y el 100% de las calificaciones de indicadores/dimensiones pudieron compararse con

¹⁴ En el anexo E se ilustra el porcentaje de indicadores y dimensiones que pueden compararse con confianza entre distintas EC cuando se eliminan los casos “sin calificar”.

confianza con las anteriores (véase el gráfico 1)¹⁵. Otras nueve EC muestran un grado de comparabilidad comprendido entre 70% y 80%. En cifras combinadas, las 24 EC representan el 73% de las EC consideradas en el presente informe.

57. Siete EC excluidas muestran un bajo porcentaje de indicadores o dimensiones que puedan compararse con confianza, es decir que menos del 50% de las calificaciones podrían compararse con calificaciones anteriores, lo que principalmente obedece a que no se adjudicó una calificación a alguna de las evaluaciones, o a ambas.

Gráfico 1: Porcentaje de indicadores o dimensiones que pueden compararse con confianza entre distintas evaluaciones comparativas, por evaluación (incluidos los casos “sin calificar”)



33 comparative assessment observations	Observaciones sobre 33 evaluaciones comparativas
Average	Promedio
Percentage of total comparable	Porcentaje del total de

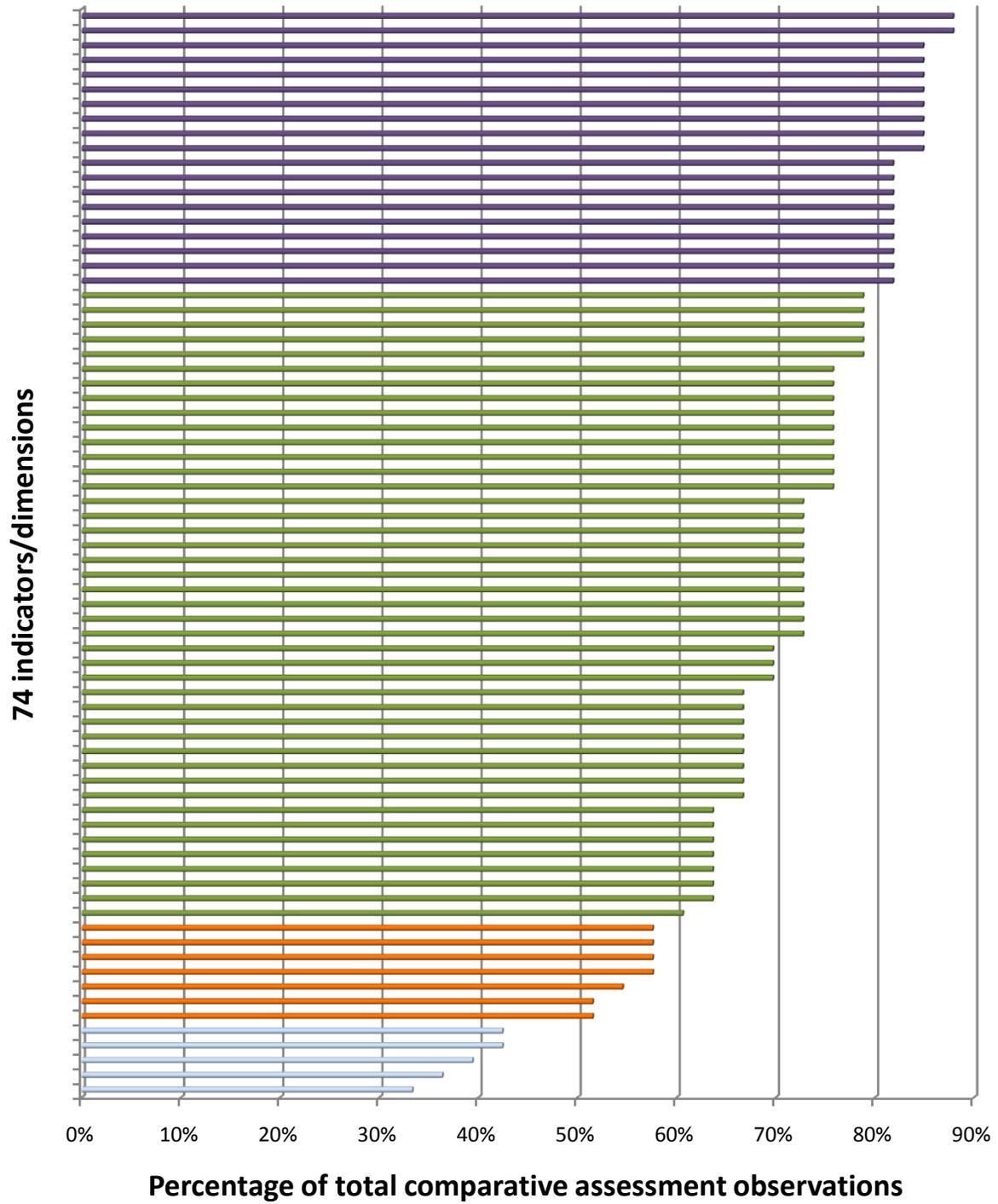
¹⁵ Entre esas 15 EC figuraron dos realizadas después de que se publicó la guía para las evaluaciones periódicas, en febrero de 2010.

indicators/dimensions	indicadores/dimensiones comparables
-----------------------	-------------------------------------

Comparabilidad entre indicadores o dimensiones (análisis horizontal)

58. Si se examinan las 33 EC, los datos indican que en promedio el 70% de los indicadores o dimensiones pueden compararse con la EA respectiva. Alrededor del 30% de los indicadores o dimensiones no pudieron compararse con confianza por falta de pruebas, por casos sin calificar en alguna de las dos evaluaciones, porque se cuestionaban las calificaciones anteriores y por otros problemas de comparabilidad. Dado que la mayoría de las EA se realizaron en los primeros dos años de la aplicación del marco, cuando la calidad de la evaluación era significativamente menor que la actual, esos resultados se consideran satisfactorios. Cabría prever que en futuros procesos de este tipo se alcanzaran tasas de comparabilidad más altas.

Gráfico 2: Porcentaje de calificaciones que pueden compararse con confianza entre distintas evaluaciones comparativas para cada dimensión del indicador (incluidos los casos “sin calificar”)



Leyenda¹⁶:

100%-80%	ID-10, ID-14 ii), ID-11 i), ID-11 iii), ID-13 ii), ID-16 ii), ID-26 iii), ID-28 i), ID-28 ii), ID-28 iii), ID-6, ID-11 ii), ID-12 iii), ID-14 iii), ID-22 i), ID-22 ii), ID-24 ii), ID-24 iii), ID-25 ii)
79%-60%	ID-14 i), ID-16 i), ID-18 i), ID-21 ii), ID-23, ID-3, ID-5, ID-12 i), ID-15 ii), ID-21 i), ID-21 iii), ID-24 i), ID-25 i), ID-26 i), ID-12 ii), ID-12 iv), ID-13 iii), ID-17 i), ID-17 ii), ID-17 iii), ID-18 ii), ID-18 iii), ID-20 iii), ID-27 iii), ID-13 i), ID-25 iii), ID-27 v), ID-2, ID-16 iii), ID-8 iii), ID-15 iii), ID-18 iv), ID-19 i), ID-19 iii), ID-20 i), ID-1, ID-4 ii), ID-8 ii), ID-9 i), ID-26 ii), ID-27 i), ID-27 ii), ID-20 ii)
59%-50%	ID-19 ii), D-2 ii), ID-7 ii), D-3, D-2 i), ID-8 i), ID-9 ii)
≤ 49%	ID-7 i), D-1 i), ID-15 i), ID-4 i), D-1 ii)

74 indicators/dimensions	74 indicadores/dimensiones
Percentage of total comparative assessment observations	Porcentaje del total de observaciones de evaluaciones comparativas

Indicadores o dimensiones con porcentajes altos/bajos de comparaciones con confianza

59. Diecinueve indicadores o dimensiones muestran un porcentaje de comparabilidad con la EA comprendido entre 80% y 100%. Por otra parte, cinco indicadores o dimensiones muestran una comparabilidad inferior al 50%. Un indicador, previsibilidad de transferencias de un más alto nivel de gobierno (HLG-1), solo se aplicó en tres EC (tres evaluaciones de Gobiernos subnacionales); este indicador no existía cuando se realizó la evaluación inicial. En consecuencia la comparación nunca sería posible, lo que implica una falta de comparabilidad del 100% y en consecuencia el caso se eliminó del gráfico para evitar sesgos en la conclusión. El gráfico 2 ilustra los porcentajes de indicadores/dimensiones (excluidas las dimensiones de HLG-1) que pueden compararse con confianza entre las 33 EC, por indicador/dimensión.

60. Los cinco indicadores o dimensiones que mostraban una comparabilidad inferior al 50% con respecto a la EA son los siguientes:

- | |
|---|
| <ul style="list-style-type: none"> • ID-4 i), Saldo y seguimiento de los atrasos de pagos de gastos • ID-7 i), Magnitud de las operaciones no declaradas del Gobierno • ID-15 i), Eficacia en materia de recaudación de impuestos • D-1 i) y ii), Previsibilidad del apoyo presupuestario directo |
|---|

61. Esos indicadores o dimensiones requieren datos cuantificables, como lo exige la metodología del PEFA, lo que no ocurre en el caso de los 19 indicadores/dimensiones altamente comparables, salvo para el ID-12 iii) (que requiere la determinación de costos correspondientes a estrategias sectoriales). Este problema podría estar vinculado con disponibilidad de datos (provenientes de las autoridades y de asociados para el desarrollo), y no necesariamente con la posibilidad de medir cambios en el desempeño basados en el marco. Un examen detallado muestra que los cinco indicadores o dimensiones poseían considerables niveles de casos “sin calificar” en alguna de las dos evaluaciones, lo que impedía la comparación. Otros 14 indicadores o dimensiones

¹⁶ Los indicadores se presentan en orden descendente. ID-10, por ejemplo, es el más próximo al 100%, en tanto que ID-25 ii) es el más próximo al 80%.

muestran más de un 15% de casos “sin calificar”¹⁷. Suponiendo que en los primeros años fuera difícil medir los cambios en el desempeño de la GFP, hay razones para creer que la falta de comparabilidad se irá reduciendo a medida que aumente la capacidad de los Gobiernos de producir la información necesaria e implementar reformas. No obstante, debería realizarse un análisis adicional para establecer si las dificultades en materia de comparación hacen necesario i) introducir cambios en los requisitos mínimos para la calificación de la dimensión; ii) establecer mejores “aclaraciones” para la aplicación del marco, o iii) establecer si una preparación insuficiente de las misiones del PEFA puede influir en la comparabilidad, ya que quizá no sea fácil obtener datos.

62. En el gráfico del anexo F se presenta el porcentaje de comparabilidad entre los indicadores cuando se excluye el factor “sin calificar”: en él se muestra que el 75% del total de los indicadores o dimensiones está comprendido dentro de la gama del 80% al 100% de comparabilidad, y el 25% en la gama del 60% al 80%; no existen indicadores o dimensiones con una comparabilidad inferior al 60%. Se excluye, como en el gráfico 2, el indicador subnacional de evaluación HLG- 1, que no figuró en ninguna EA.

“Problemas de seguimiento”

63. Se identificaron cinco “problemas de seguimiento” que impidieron identificar la comparación: a) nuevas pruebas no disponibles (o no utilizadas) en la EA; b) cambios de definiciones; c) diferentes interpretaciones de los requisitos del marco; d) diferentes bases de toma de muestras, y e) el hecho de que en las EC se haya modificado la calificación de la EA para corregir errores tangibles. Corresponde subrayar que ello representa un enfoque sumamente prudente para poder establecer comparaciones. Por ejemplo, cuando el equipo encontró casos de EC en que se recalcularon calificaciones anteriores para identificar cambios, no se identificó la orientación del cambio a lo largo del tiempo, dada la imposibilidad de establecer comparaciones. En la guía para las evaluaciones periódicas (febrero de 2010) se indica que no corresponde volver a calificar errores irrefutables existentes en las EA, pero que ellos deben declararse como parte del análisis de los cambios¹⁸.

64. Los datos emanados de 33 EC realizadas mediante 2451 observaciones, muestran que se identificaron problemas de “seguimiento” 175 veces; de ellos, i) el 24% guardaba relación con nuevas pruebas recogidas con respecto a la EA; ii) el 14% eran casos de cambios de definiciones; iii) el 45% se refería a diferencias en la interpretación; iv) el 6% guardaba relación con diferencias en el muestreo, y v) el 11% correspondía a casos en que la EC recalificó la EA. Algunos “problemas de seguimiento” pueden evitarse con mayor facilidad que otros (diferencias en la interpretación, diferencias en el muestreo y recalificación de la EA, en contraposición con nuevas pruebas recogidas con respecto a la EA y cambios de definiciones). Los problemas más fácilmente evitables representan el 62% del total de casos; no obstante, en tanto que las “diferencias en la interpretación” constituyen un problema específico de los evaluadores, los otros dos pueden considerarse en mayor medida como problemas de proceso. Debería realizarse una labor de seguimiento para identificar las dimensiones más propensas a problemas de seguimiento, así como las razones que explican la incidencia de problemas de ese tipo, y recomendar

¹⁷ ID-4 ii), ID-7 ii), ID-8 (i, ii, iii), ID-9 (i, ii), ID-12 ii), ID-15 iii), ID-17 (i, iii), D-2 (i, ii), D-3.

¹⁸ Téngase presente que no será necesario modificar las calificaciones de tres indicadores en relación con los cuales se modificaron recientemente las escalas de calificación. En los casos de ID-2 i), ID-3 e ID-19 iii) los equipos de evaluación tendrán que volver a calcular la calificación anterior utilizando los nuevos criterios para identificar la orientación del cambio a lo largo del tiempo.

las medidas que deben adoptarse. En el recuadro 1 aparecen ejemplos de “*problemas de seguimiento*”.

Posibles factores de procesos que afectan a la calidad de la medición de los cambios

65. Para examinar esos factores se consideraron las notas conceptuales y los términos de referencia, el uso del mismo equipo de evaluación en la EA y en la EC, el nivel de participación del Gobierno en la evaluación, el hecho de que se hubieran organizado o no talleres iniciales y finales, y el proceso de examen del control de calidad.

Recuadro 1. Ejemplos de “problemas de seguimiento”

Nuevas pruebas recogidas con respecto a la EA: la evaluación actual permitió obtener información que no estaba disponible en la EA.

Cambios de definiciones: por ejemplo las de atrasos, atrasos internos, fondos extrapresupuestarios, clasificación de entidades paraestatales, entidad subnacional versus entidad desconcentrada del Gobierno central.

Diferencias en la interpretación: por ejemplo: ID-24 ii): en la EC se constató que los informes de ejecución del presupuesto del ejercicio anterior se habían publicado entre cuatro y siete semanas después de finalizado el trimestre, con un promedio de 5,5 semanas, y se asignó una calificación “B”. En la EA se constató que el desempeño fue exactamente el mismo, pero se otorgó una calificación “C”, debido a que dos informes se retrasaron más de seis semanas.

Diferencias en el muestreo: por ejemplo. ID-21: la EA se basó en información sobre la situación de los ministerios de Agricultura, Justicia y Finanzas. La EC se basó en información de los ministerios de Educación y Justicia. ¿Es válida una comparación directa?

Evaluación comparativa que modifica la calificación de la EA: por ejemplo: i) otorgar una calificación errónea a pesar de que las pruebas son muy claras; ii) otorgar “+” a un indicador de una sola dimensión.

66. *Claridad de las notas conceptuales y los términos de referencia con respecto a la medición de los cambios.* Se analizaron las notas conceptuales y los términos de referencia de 31 EC; evaluaciones del Gobierno federal de Etiopía y de tres Gobiernos subnacionales de ese país quedaron comprendidas en la misma nota conceptual/los mismos términos de referencia, pero con fines estadísticos se las consideró como cuatro notas conceptuales/términos de referencia. En dos EC no se prepararon notas conceptuales/términos de referencia.

67. Solicitar comentarios a la Secretaría del PEFA sobre el borrador de las notas conceptuales y los términos de referencia dista mucho de ser la práctica actual. Al iniciarse la preparación de este informe de seguimiento menos de la mitad de las notas conceptuales y los términos de referencia correspondientes a las evaluaciones consideradas en el presente informe estaban disponibles en la secretaría. Se obtuvieron ulteriormente de los interesados, de informes finales y de consultores. En general la secretaría no recibe la versión final después de haber emitido sus comentarios; en algunos casos las notas conceptuales y los términos de referencia se transmiten a la secretaría pero con excesiva tardanza como para que pueda proporcionar comentarios.

68. El análisis reveló que en 24 notas conceptuales/términos de referencia (el 73%) era necesario medir los cambios en el desempeño registrados después de la EA del PEFA. La manera en que se expresa ese requisito se divide principalmente en dos categorías: 1) una *referencia simple* a “medición del progreso en materia de GFP después de la última evaluación”, y 2) una *clara* descripción de lo que se espera de la EC.

69. En 14 notas conceptuales y términos de referencia de las 15 EC con un 80% a 100% de indicadores o dimensiones comparados con confianza se disponía la medición de los cambios experimentados desde la EA, y nada se decía en la restante. De las siete EC con un bajo porcentaje de comparabilidad de indicadores o dimensiones (menos del 50%), en cuatro notas conceptuales y términos de referencia se solicitaba la medición de cambios en el desempeño (tres evaluaciones de Gobiernos subnacionales estaban cubiertas por una misma evaluación del Gobierno central), en dos casos nada se decía, y en uno no se había preparado¹⁹.

70. Lo expresado confirma la importancia de “asegurarse de que la nota conceptual/los términos de referencia sean lo suficientemente minuciosos, y de que todas las partes interesadas los entiendan claramente”, como lo recomendaba la Secretaría del PEFA en la nota sobre *Prácticas recomendadas para las evaluaciones periódicas*.

71. Las pruebas anecdóticas indican que en las notas conceptuales y los términos de referencia de las ER debería preverse el tiempo suficiente para verificar los fundamentos de las calificaciones anteriores adjudicadas. Aunque esto puede considerarse como la adjudicación implícita de una calificación al llevar a cabo una ER, no otorgar el tiempo suficiente para esa tarea puede contribuir a impedir o limitar un análisis profundo de la EA.

72. *Uso del mismo equipo de evaluación en ambas evaluaciones.* La composición del equipo de evaluación en evaluaciones anteriores y comparativas muestra que en nueve casos el jefe de equipo era el mismo o algunos miembros del equipo habían tomado parte en ambas evaluaciones (en dos casos el jefe del equipo de la EA era un miembro del equipo de la EC). El análisis muestra que esos nueve casos estaban comprendidos dentro las 15 EC con altos porcentajes de indicadores o dimensiones comparados con confianza. En ninguna de las siete EC con bajos porcentajes de comparación de indicadores o dimensiones existía coincidencia de miembros de equipos de evaluación²⁰. El uso de los mismos evaluadores fue considerado en tres países como “una indudable ventaja”, “los evaluadores se habían ganado la confianza del Gobierno y de los asociados en el desarrollo” y el “Gobierno confiaba en los consultores que habían actuado anteriormente” (en el borrador de noviembre de 2010 de *Assessing the Impact of the PEFA Framework*). Ello pone de manifiesto un factor provechoso en el proceso de medición del progreso a lo largo del tiempo, pero suele no ser posible utilizar el mismo equipo de evaluación para la EC. Por otra parte, aunque el uso del mismo equipo de evaluación puede contribuir a reducir el número de cambios espurios producidos por diferentes juicios o interpretaciones (o incluso diferentes sesgos), al mismo equipo puede resultarle más difícil aceptar que la EA hubiera sido insatisfactoria en general²¹. En otros términos, utilizar el mismo equipo de evaluación solo es ventajoso si la EA se considera de buena calidad.

¹⁹ De las 12 ER no incluidas en el presente estudio solo en una la nota conceptual y los términos de referencia requerían la medición de cambios en el desempeño a lo largo del tiempo (Tres notas conceptuales/términos de referencia no estaban disponibles; en uno de esos casos se trataba de una autoevaluación).

²⁰ Ninguna de las 12 ER descalificadas para el presente estudio había contado con el mismo jefe de equipo/los mismos evaluadores en ambas evaluaciones.

²¹ No obstante, la experiencia muestra que los evaluadores están dispuestos a admitir la posible existencia de unas pocas fallas en su evaluación anterior y por lo tanto la necesidad de reconsiderar las calificaciones.

73. En los casos en que no es posible garantizar que algunos de los evaluadores participe en la ER cabe la posibilidad de establecer contacto con el jefe de equipo de la EA para que el equipo de la ER pueda obtener información detallada (por ejemplo a partir de notas de los evaluadores) basada en las calificaciones de la EA, y analizar la validez de los cambios en el desempeño identificados. La secretaría ha participado en unos pocos casos de este género, lo que ayudó a los evaluadores actuales a conocer los niveles de desempeño anteriores y abstenerse de formular conclusiones sobre cambios de calificaciones basadas en supuestos imprecisos, lo que reviste especial importancia cuando los evaluadores de ER reciben información que pone en tela de juicio la validez de las calificaciones de la EA.

74. *Participación del Gobierno en la evaluación comparativa.* Pueden distinguirse dos tipos de participación del Gobierno en la preparación de la EC: i) “*cooperación y respaldo para el equipo de evaluación*”, que representa un papel más pasivo, y ii) “*orientación del proceso y participación profunda en el mismo*”, que supone un papel más activo. Doce países —en lo que se considera una muy buena práctica— establecieron comités de orientación del PEFA o estructuras similares (que variaron de un país a otro) para realizar el seguimiento, la supervisión y la revisión de la evaluación. Otro factor que puede haber influido en la calidad de la EC es la organización de talleres iniciales y de cierre: en tanto que los primeros contribuyen a dar a conocer mejor a los funcionarios la metodología del PEFA y su participación en el proceso del PEFA, el seminario de cierre ofrece un foro para analizar las conclusiones y recibir comentarios. Dieciséis EC organizaron talleres iniciales y finales²².

75. De las 15 EC con 80% a 100% de indicadores o dimensiones comparados con confianza, cinco establecieron una estructura específica y organizaron talleres iniciales y de salida; salvo en un caso, esos países no coinciden. Por el contrario ninguna de las siete EC con menos de 50% de indicadores o dimensiones comparados con confianza estableció una estructura específica para el seguimiento del proceso²³ y solo una organizó ambos talleres.

76. *Examen de garantía de calidad.* Todos los borradores de los informes de ER (incluidos los 12 que no figuran en el presente estudio) fueron enviados a la Secretaría del PEFA para que formulara comentarios, con excepción de dos autoevaluaciones (la de Uganda de 2008 y la de la República Dominicana de 2009) (que no eran EC). El examen de calidad llevado a cabo por la Secretaría del PEFA podría considerarse como un elemento importante para proporcionar a los interesados garantía sobre la calidad de la evaluación.

77. **En resumen,** los factores del proceso que anteceden pueden haber contribuido a dar solidez a la medición de los cambios a lo largo del tiempo, tal como se señala en el

²² No en todos los informes se proporciona sistemáticamente información referente al grado de participación del Gobierno y a los talleres; en algunos casos no es suficientemente específica como para que abarque la medida en que las autoridades hayan tomado parte en el proceso, por lo cual esa cifra probablemente será más alta (por ejemplo, la información disponible muestra que se organizaron 20 talleres iniciales y 19 finales).

²³ Los datos de las 12 ER no incluidas en el presente estudio muestran que en dos casos las autoridades han establecido una estructura para la gestión de la evaluación del PEFA.

cuadro 4, *infra*; la incidencia de atributos positivos en el proceso de evaluación parece haber contribuido a elevar el nivel de comparabilidad, lo que no parece suceder cuando el proceso de evaluación presenta, a lo sumo, escasos componentes generadores de “buenos procesos”. El cuadro 4 ilustra las observaciones de los factores del proceso de las 33 EC y los que tienen grados de comparabilidad más altos y más bajos; los porcentajes se calculan en relación con el total de las EC, el número de las EC comparables en una gama del 80% al 100% y el número total de EC con menos de 50% de comparabilidad. Como ya se mencionó, los informes no siempre contienen información referente a la participación del Gobierno y a los talleres, y cuando la hay, suele ser vaga. Es más probable que su incidencia sea mayor que la información actualmente disponible. Con el tiempo, cuando se disponga de más evaluaciones periódicas, sería útil llevar a cabo una labor de seguimiento para tener idea de cuál de esos factores es más importante a los efectos de la comparabilidad.

78. De los factores de proceso considerados, solo el examen por la Secretaría del PEFA de los borradores de los informes no está correlacionado con el grado de comparabilidad, por la sencilla razón de que todas las EC comprendidas en el análisis habían sido examinadas por la secretaría. No obstante, anteriores informes de seguimiento han demostrado que los exámenes de la secretaría (así como los llevados a cabo por otros interesados) contribuyen significativamente a mejorar la tasa de cumplimiento y por lo tanto la medida en que las calificaciones de los indicadores cuentan con pruebas adecuadas. En consecuencia, los exámenes realizados por la secretaría influyen sobre la solidez de la comparación de las calificaciones de las EA y las EC.

79. Tal como se reconoce en el documento de la Secretaría del PEFA “Buenas prácticas en la aplicación del marco de referencia para la medición del desempeño en materia de GFP”, el proceso de evaluación *“tiene que estar bien gestionado a fin de contribuir a garantizar un producto de alta calidad ... cuanto más destacado sea [el papel que el Gobierno desempeñe] mayor será el beneficio que obtendrá el personal gubernamental del ejercicio ... para que los Gobiernos sean capaces de ejercer una sólida función de liderazgo es fundamental que comprendan la metodología del marco y el proceso de realización de una evaluación PEFA”*. No menos importante es el hecho de que por más que exista un proceso adecuadamente diseñado y aplicado, el mismo puede dar lugar a sucesivas evaluaciones, actualizaciones y revisiones del PEFA que hagan del seguimiento de los cambios en el desempeño a lo largo del tiempo una tarea difícil y aumenten las probabilidades de pasar por alto cambios en el desempeño que se hayan producido entre tanto (por ejemplo, en Tanzania, tras varias actualizaciones y revisiones realizadas desde la evaluación del PEFA de 2005, la evaluación de 2010 es la primera en que se intentó explícitamente medir el progreso).

Cuadro 4: Incidencia de factores de proceso en las evaluaciones comparativas

Factores de proceso	Todas las EC	%	EC comparables 80%-100%	%	EC comparables ≤ 50%	%
---------------------	--------------	---	-------------------------	---	----------------------	---

Total de EC	33	100	15	100	7	100
Claridad de las notas conceptuales y los términos de referencia	24	73	14	93	4*	57
Uso del mismo equipo de evaluación	9	27	9	60	0	0
Participación del Gobierno (estructura específica)	12	36	5	33	0	0
Talleres inicial y de salida	16	48	5	33	1	14
Examen de garantía de calidad	33	100	15	100	7	100

*Incluye tres evaluaciones de Gobiernos subnacionales realizadas junto con una evaluación del Gobierno federal.

80. En conclusión, al examinar 33 EC se observa que la comparabilidad se ve afectada por la proporción de indicadores o dimensiones “sin calificar” en alguna de las dos evaluaciones, o en ambas, que es mayor en los casos en que las pruebas son inadecuadas. El grado de comparabilidad aumenta cuando se excluye del análisis el factor “sin calificar”, lo que se aplica tanto a los análisis verticales como a los horizontales. Los “problemas de seguimiento” que impidieron la comparación guardaban relación, principalmente, con diferencias de interpretación y con nuevas pruebas recogidas con respecto a la EA. Además algunos factores de proceso pueden haber influido en la calidad de la medición de los cambios a lo largo del tiempo. La presencia de esos factores parece haber contribuido a elevar el nivel de comparabilidad, pero no se observa otro tanto cuando su presencia es limitada, como se comprueba cuando se realiza un examen cruzado entre EC con grados de comparabilidad altos y bajos.

Capítulo 4: Tendencias en el desempeño en materia de gestión de las finanzas públicas

4.1 Reseña

81. En esta sección se examinan las siguientes cuestiones:

- ¿Qué indicadores están cambiando?
- ¿Existen modalidades de cambio entre diferentes tipos de indicadores y características de países? ¿Y entre regiones diferentes?
- ¿Cuáles son los posibles factores explicativos?

4.2 Conclusiones

82. Se analizaron modalidades de cambio para todos los indicadores o dimensiones en que puede realizarse una comparación válida del desempeño entre dos evaluaciones consecutivas. Por lo tanto, se cree que esos cambios representan cambios reales en el desempeño, y no simplemente nuevas calificaciones basadas en mejor información o diferentes juicios o interpretaciones. De todas las calificaciones de indicadores o dimensiones, el 11% mantuvieron calificaciones “A”, el 21% mejoraron, el 10% mantuvieron calificaciones “D”, el 9% empeoraron, el 21% mantuvieron calificaciones “B” o “C”, y el 30%²⁴ de los casos no se prestaban a una medición válida del cambio. Los indicadores o las dimensiones siguientes son las que registraban los mejores y peores desempeños:

- El ID-24 ii), *Puntualidad de la presentación de los informes presupuestarios del ejercicio en curso*, obtuvo el mayor porcentaje (48%) de calificaciones “A” en dos evaluaciones sucesivas.
- Los más altos porcentajes siguientes correspondieron al ID-3, *Desviaciones de los ingresos agregados en comparación con el presupuesto original aprobado* (45%) y al ID-15 ii), *Eficacia de la transferencia del cobro de tributos en la Tesorería por parte de la administración de ingreso fiscal* (36%).
- Dos dimensiones, ID-8 iii) *Medida en que se recogen y dan a conocer datos fiscales consolidados (por lo menos sobre ingresos y gastos) referentes al Gobierno general, por categorías sectoriales*, e ID-28 i), *Puntualidad del examen de los informes de auditoría por parte del Poder Legislativo (en el caso de informes recibidos dentro de los tres últimos años)* fueron las que obtuvieron el mayor porcentaje (39%) de calificaciones “D” en dos evaluaciones sucesivas.
- El más alto porcentaje siguiente de calificaciones “D” correspondió al ID-23 *Disponibilidad de información sobre los recursos recibidos por las unidades de prestación de servicios* (33%).
- El ID-13 ii), *Acceso de los contribuyentes a información sobre responsabilidades y procedimientos administrativos en materia tributaria* registró el más alto porcentaje de casos con aumento de las calificaciones en dos evaluaciones sucesivas (48%).
- Tres indicadores o dimensiones obtuvieron el más alto porcentaje de casos con aumento de las calificaciones en dos evaluaciones sucesivas: ID-14 iii),

²⁴ La suma supera el 100% debido al redondeo de las cifras.

- Planificación y control de los programas de auditoría e investigaciones de fraude tributario (42%); ID-12 ii), Alcance y frecuencia de los análisis de sostenibilidad de la deuda, e ID-14 ii), Eficacia de las sanciones por incumplimiento de obligaciones de registro y de declaración (38%).*
- Cuatro indicadores o dimensiones obtuvieron el más alto porcentaje de casos con reducción de las calificaciones en dos evaluaciones sucesivas: ID-23, *Disponibilidad de información sobre los recursos recibidos por las unidades de prestación de servicios (21%); ID-1, Resultados del gasto agregado en comparación con el presupuesto original aprobado; ID-3, Desviaciones de los ingresos agregados en comparación con el presupuesto original aprobado, e ID-17 ii), Medida de la consolidación de los saldos de caja del Gobierno (18%).*

83. Más países obtuvieron un mayor número de calificaciones máximas o mejoradas (23 países) que de calificaciones mínimas o menos satisfactorias (ocho países)²⁵, lo que indica una tendencia amplia y alentadora de mejora de la GFP en los distintos países examinados. Cinco de estos últimos eran pequeños Estados insulares con dificultades que no son típicas de la muestra en general, incluida una gran vulnerabilidad frente a conmociones externas tales como huracanes e inestabilidad de los ingresos del turismo, graves restricciones de capacidad con intensa emigración de recursos humanos calificados, e inestabilidad política. Los otros tres eran Estados africanos que enfrentaban muchos de esos desafíos.

84. Es posible analizar en mayor detalle las modalidades de desempeño globales utilizando una metodología (Andrews, 2009; Porter y otros, 2010) en que los indicadores o dimensiones se clasifican en tres parejas (el análisis “solo considera los indicadores/dimensiones ID-5 a ID-28, ya que los indicadores ID-1 a ID-4 abarcan resultados y desempeño del sistema de GFP, y no la calidad de los sistemas de GFP de por sí”)²⁶.

85. En la primera pareja se contraponen las dimensiones del PEFA en que puede obtenerse un puntaje “C” o superior, a través de una nueva ley o el anuncio de una nueva práctica, aunque esta no se implemente (*de jure*) con dimensiones que requieran la implementación actual o un compromiso significativo (*de facto*)²⁷. Por ejemplo, en el caso de la dimensión *de jure* ID-11 i) se obtiene una calificación “C” en la medida en que exista un calendario presupuestario anual, aunque pueden existir considerables demoras de implementación, sin que se conceda tiempo suficiente para que las entidades presupuestarias completen estimaciones detalladas. Por otra parte, la dimensión *de facto* ID-12 i) requiere la producción efectiva, con carácter anual, de previsiones de los agregados fiscales para los dos años siguientes.

86. En la segunda pareja se contraponen las dimensiones del PEFA relativas a preparación presupuestaria, tales como presupuestación estratégica (previsiones plurianuales, planificación estratégica, planificación de inversiones, planificación de la

²⁵ En dos países el número de calificaciones máximas o mejoradas equivalió al de calificaciones mínimas o menos satisfactorias. Las calificaciones máximas o mejoradas se combinan porque no es posible mejorar una calificación “A”; las calificaciones mínimas o menos satisfactorias se combinan porque no es posible que una calificación sea inferior a “D”.

²⁶ En el anexo H aparece la lista completa de los indicadores/dimensiones codificados.

²⁷ Los indicadores/dimensiones fueron codificados en forma independiente por tres especialistas en GFP; las discrepancias se analizaron y reconciliaron (Andrews, 2009). En el anexo H aparecen todas las codificaciones.

deuda); preparación del presupuesto anual; análisis parlamentario del presupuesto anual, y la estructura de documentos presupuestarios formales por una parte (**iniciales**), y dimensiones relativas a la ejecución presupuestaria tales como gestión de recursos (incluidas gestión de entradas y salidas de efectivo, adquisiciones, nómina); control interno, auditoría y seguimiento internos, contabilidad e información, auditoría externa y análisis parlamentario de informes de auditoría, por otra parte (**ulteriores**). Las primeras tienen que ver con la etapa inicial del ciclo de planificación presupuestaria, visible para los donantes y los inversionistas, y cabría prever que tropiecen con menos resistencia que los aspectos mencionados en segundo término, referentes a controles y supervisión del gasto real. Por ejemplo, en el caso del indicador inicial ID-5 se obtiene una calificación “C” si el presupuesto se clasifica utilizando las estadísticas de las finanzas públicas o estándares similares: se trata de una práctica formal que puede ser objeto de seguimiento por donantes e inversionistas. Por otra parte, los elementos ulteriores tienen que ver con cuestiones más delicadas de gestión y seguimiento de gastos reales. Por ejemplo, en el caso del indicador ulterior ID-7, una calificación “C” requiere que el gasto presupuestario no declarado no pase del 5% al 10% del total del gasto. Ello reduce la posibilidad de encubrimiento de fondos, en forma no transparente, que en muchos países superan ampliamente esa proporción, y que pueden tropezar con fuerte resistencia de intereses influyentes beneficiados por tales sistemas.

87. En la tercera pareja se contraponen aspectos del PEFA sujetos al control de órganos centrales, reguladores, como el Ministerio de Finanzas (**control concentrado**), con aquellos que requieren la participación de varios organismos o autoridades subnacionales (**control desconcentrado**). En el caso de la dimensión concentrada ID-12 ii), por ejemplo, se obtiene una calificación “C” en la medida en que se haya realizado en los últimos tres años un análisis de sostenibilidad de la deuda por lo menos para la deuda externa, labor técnica que puede realizar un pequeño equipo de funcionarios técnicos. Por otra parte, la dimensión desconcentrada ID-12 iii) requiere el cálculo de los costos de las estrategias sectoriales correspondientes a varios grandes sectores, lo que exige la participación de varias entidades presupuestarias.

88. Aunque existe superposición entre las tres parejas, ellas ilustran en forma amplia la contraposición entre forma y función. Como norma general, los aspectos *de jure*, iniciales y concentrados abarcan los atributos formales, en tanto que los aspectos *de facto*, ulteriores y desconcentrados abarcan los atributos funcionales de un sistema. El análisis estático de calificaciones del PEFA realizado por investigadores independientes indica que las calificaciones no inferiores a “C” adjudicadas a los atributos formales se logran más comúnmente que esas mismas calificaciones para atributos funcionales. La explicación que se presume es que el progreso formal puede lograrse a través de la adopción de nuevas leyes, reglamentos o instrumentos técnicos o la participación de tan solo unos pocos organismos, o en una etapa temprana del ciclo presupuestario; más difícil es lograr progreso funcional, porque coordinar la labor de numerosos organismos es una labor más ardua y es más difícil realizar el seguimiento de reformas efectivamente implementadas, que pueden comprometer la obtención de rentas y suscitar una más intensa oposición política y burocrática²⁸.

89. En el mismo orden de ideas cabría prever una mayor propensión de altas calificaciones para los atributos formales, o de bajas calificaciones, o deterioro, para los

²⁸ Los términos “formal” y “funcional” se refieren a la tendencia global de un indicador o dimensión, en algunos de los cuales se combinan atributos formales y funcionales. Esas denominaciones no constituyen recomendaciones con respecto a la secuencia de las reformas, ya que puede ser necesario que las reformas funcionales se inicien tempranamente.

atributos funcionales. Esa es, en efecto, en términos amplios, la modalidad que aquí se observa. Es mayor la probabilidad de que los atributos formales mejoren o se ubiquen en el máximo nivel, y que los funcionales se reduzcan o se ubiquen en el nivel mínimo. Son muchas las calificaciones en proceso de mejoramiento y de máximo nivel en las esferas funcionales, lo que demuestra que incluso en la esferas de mayores dificultades es posible obtener calificaciones altas o que vayan mejorando. No obstante, el alto número de calificaciones mínimas o en pronunciado proceso de deterioro en esferas funcionales lleva a pensar que el avance en esas esferas puede ser frágil, y presentar riesgos de retroceso.

90. Esta metodología puede utilizarse para contribuir a explicar el desempeño antes indicado en materia de calificaciones. Los tres indicadores/dimensiones que mantenían calificaciones “D” en el más alto porcentaje de casos constituían ejemplos de las tres características funcionales (*de facto*, desconcentradas y ulteriores), en que cabría prever un progreso relativamente arduo. En el caso de los dos indicadores/dimensiones que indican la calidad de los sistemas de GFP que empeoraron en el más alto porcentaje de casos, uno (ID-23) fue asimismo un ejemplo de los tres atributos funcionales, en tanto que el otro (ID-17 ii) representó un caso en que las dificultades de una labor en una etapa ulterior del ciclo presupuestario superaron abrumadoramente a los beneficios que suponía el que esos atributos fueran concentrados y *de jure*. Las dos dimensiones que mantenían calificaciones “A” en el máximo porcentaje de los casos, y cuatro que mejoraron en el máximo porcentaje de los casos, fueron ejemplos en que el control centralizado estuvo a cargo de órganos de ejecución, en los que cabía esperar éxito en etapas tempranas, ventaja esta suficiente para compensar los efectos de los componentes *de facto* de algunos de esos atributos, que presentan mayores dificultades.

91. Cabe discernir modalidades similares en otros aspectos del desempeño. Por ejemplo, el ID-25 ii), *Puntualidad en la presentación de los estados financieros*, presentó el máximo porcentaje de mejoramiento de las calificaciones, que pasaron de “D” a “A” (15%), y está sujeto al control centralizado de un órgano de ejecución, como ocurre con el ID-13 ii), *Acceso de los contribuyentes a información sobre responsabilidades y procedimientos administrativos en materia tributaria*, al que correspondió el máximo porcentaje de casos que avanzaban de “C” a “A” (12%), y con el ID-14 iii), *Planificación y control de los programas de auditoría e investigaciones de fraude tributario*, en que se registró el máximo porcentaje de casos que pasaban de “C” a “B” (24%). Esas dimensiones de desempeño adecuado tenían también cierto grado de correspondencia con atributos *de jure* e iniciales, aunque en forma menos sistemática. Por otra parte hubo una categoría de desempeño adecuado, en que el más alto porcentaje de casos pasaban de “D” a “C”, en que una de las dos dimensiones de desempeño más satisfactorio era un ejemplo de tres atributos funcionales (*de facto*, desconcentradas y ulteriores), ID-26 i), *Oportunidad de la presentación de informes de auditoría al poder legislativo* (18%). La otra dimensión de óptimo desempeño, ID-12 iii), *Existencia de estrategias sectoriales con determinación plurianual de costos del gasto recurrente y de inversiones* (18% en ambos casos), fue un ejemplo de dos atributos funcionales (*de facto* y desconcentradas). Esos ejemplos demuestran que determinados Gobiernos quizá puedan llevar adelante reformas más arduas si existe una oportunidad política o se cuenta con un respaldo más sólido de asociados para el desarrollo.

92. El cuadro 5 muestra un resumen del número de calificaciones máximas o en proceso de mejoramiento por tipo de dimensión. Ambos se presentan combinados, porque no es posible obtener una calificación mejor que “A”. Los atributos formales de la

izquierda representan un resultado levemente mejor que los atributos funcionales de la derecha, ya que cabía prever que los aspectos sujetos al control de entidades concentradas variarían más rápidamente que las leyes y normas formales que tendrían que haberse sancionado antes de que pudieran tener lugar mejoras funcionales²⁹. No obstante, las diferencias de porcentajes son más reducidas que las que podrían preverse, existiendo una mínima diferencia entre atributos iniciales y ulteriores, lo que indica la complejidad de cada caso en que el contexto, las intervenciones de los donantes, las redes de patrocinio y toda clase de factores de otro género se interrelacionan y afectan el desempeño. En las tres esferas funcionales de la derecha abundan las calificaciones máximas o en proceso de mejoramiento, lo que muestra que incluso en las esferas de mayores dificultades es posible obtener calificaciones altas o en proceso de mejoramiento.

Cuadro 5: Calificaciones máximas o en proceso de mejoramiento, por tipo de indicador/dimensión

Tipo de dimensión	Porcentaje de calificaciones	Tipo de dimensión	Porcentaje de calificaciones
Concentrada	41	Desconcentrada	26
<i>De jure</i>	35	<i>De facto</i>	31
Inicial	33	Ulterior	32

93. En el cuadro 6 se examinan calificaciones mínimas o en proceso de deterioro. También en este caso unas y otras se combinan, porque no es posible obtener una calificación inferior a “D”, y una modalidad similar resulta evidente si se examinan exclusivamente las calificaciones en proceso de deterioro. Los datos obtenidos muestran que es mayor la probabilidad de que los atributos funcionales de la derecha obtengan calificaciones mínimas o reveladoras de un pronunciado proceso de deterioro que los atributos formales de la izquierda. Esto puede indicar que los avances en las esferas funcionales pueden ser frágiles y entrañar riesgos de retrocesos.

94. Si se comparan los cuadros 5 y 6 se advierte que es mucho mayor la proporción de las calificaciones de atributos formales de niveles máximos o en proceso de mejoramiento que las que corresponden a calificaciones mínimas o en proceso de deterioro, lo cual era previsible. Las diferencias entre calificaciones formales y funcionales son mayores para los puntajes mínimos o en proceso de deterioro que para los máximos o en proceso de mejoramiento. En otros términos, la comparación de las calificaciones de atributos funcionales con las de los formales parecen revestir mayor importancia en el extremo inferior que en el superior del espectro. No obstante, resulta sorprendente que entre los atributos funcionales exista una proporción levemente mayor de calificaciones crecientes o más altas que calificaciones mínimas o decrecientes, lo que indica que las reformas están suscitando resultados aun en las esferas de mayores dificultades. Una posible explicación es que puede ser más fácil mejorar lo que se está haciendo de manera insatisfactoria que lo que ya está dando buenos resultados (como ocurre con muchos de los atributos formales). Una diferencia de pruebas de medios indica que esos resultados son significativos al nivel del 95%³⁰.

²⁹ Una modalidad similar resulta evidente si se limita el análisis a las calificaciones en proceso de mejoramiento.

³⁰ Sería útil agregar al conjunto de datos, en una etapa ulterior, las 12 ER no comparativas, puesto que la imposibilidad de realizar comparaciones no hace más que agravar el error de medición aleatorio. Ello ayudaría a establecer si las mismas modalidades se dan en la muestra en general.

Cuadro 6: Calificaciones mínimas o en proceso de deterioro, por tipo de indicador/dimensión

Tipo de dimensión	Porcentaje de calificaciones	Tipo de dimensión	Porcentaje de calificaciones
Concentrada	9	Desconcentrada	24
<i>De jure</i>	11	<i>De facto</i>	23
Inicial	15	Ulterior	19

95. Las calificaciones mínimas y reveladoras de un más pronunciado proceso de deterioro motivan una preocupación adicional. De los cuatro indicadores o dimensiones que mantuvieron calificaciones “D” o que implican las más pronunciadas disminuciones, y que pueden caracterizarse como procesos, todas eran *de facto* y desconcentradas, y tres de las cuatro eran ulteriores. Esto pone aún más de manifiesto las dificultades de iniciar el proceso de superación y mantener el progreso logrado anteriormente en esas arduas esferas funcionales de la GFP.

96. El equipo examinó asimismo los cambios experimentados por las calificaciones de los siete casos de intervalos breves; los resultados aparecen en el cuadro 7. Aunque se preveía que los intervalos de menos de tres a cinco años serían demasiado breves como para poner en evidencia el progreso, en realidad el número de calificaciones en proceso de mejoramiento era mayor en los casos de intervalos breves, y el número de calificaciones que no podían compararse también era más reducido. La explicación del porcentaje sorprendentemente alto de calificaciones máximas y en proceso de mejoramiento podría ser que en todos los casos de EC de intervalos breves un factor clave de motivación de la realización de la evaluación era que ello representaba una condición para el respaldo de los donantes, lo que constituye un posible incentivo para exhibir el sistema de GFP a la luz más favorable posible. También puede haber sido un incentivo para que el Gobierno y los donantes realizaran tempranamente una evaluación periódica si los interesados estaban convencidos por anticipado de que ella pondría de manifiesto una significativa mejora de los sistemas de GFP.

97. Kosovo es un ejemplo de país en que las evaluaciones se realizaron a intervalos breves y logró rápido progreso, manteniendo una calificación “A” o con mejoras en el 47% de los indicadores/dimensiones. Según el informe de la EC de ese país ello se explica, principalmente, por el rápido avance en el mejoramiento del sistema de ejecución presupuestaria a través de un sistema de información de la administración financiera, un gráfico de cuentas y una cuenta única de Tesorería. Las recientes mejoras fueron también beneficiosas para los sistemas de auditoría y control internos y auditoría externa. Otro de esos casos es el de Mozambique. Según el informe de la EC de ese país, la mayoría de las mejoras fueron orientadas por reformas en curso en materia de recaudación y gestión de ingresos, adquisiciones y gestión financiera vinculada con la implementación en curso del sistema de información sobre administración financiera de Mozambique, e-SISTAFE. Algunas de las mejoras fueron el resultado de un examen de los resultados de la EA, y muchas se encontraban en estadios avanzados de diseño cuando se llevó a cabo la EA. Esos y otros casos de intervalos breves no son típicos de la muestra en general, pero revelan la posibilidad de lograr un rápido progreso en la consecución de ciertas mejoras, en contextos propicios.

Cuadro 7: Resultados de EC de intervalos breves en comparación con promedios globales

	Porcentaje medio de países con intervalos breves	Porcentaje medio global
Calificaciones “A” mantenidas	8	11
Calificaciones crecientes	31	20
Calificaciones “D” mantenidas	8	10
Calificaciones decrecientes	8	8
Calificaciones “B” o “C” mantenidas	22	21
Calificaciones no comparables	24	30

98. Finalmente, el equipo examinó posibles correlaciones entre medición del desempeño y características de los países. En un estudio anterior (de Renzio 2009) se utilizaron análisis de regresión múltiple para comparar calificaciones estáticas del PEFA con características de los países. Se concluyó que las calificaciones de la GFP están asociadas positivamente con el producto interno bruto (PIB) per cápita, el crecimiento económico reciente y la democracia, y asociadas negativamente con los recursos naturales y la dependencia de la ayuda, todo lo cual confirma las previsiones.

99. Con respecto a la medición del desempeño también cabría prever asociaciones negativas entre, por una parte, mantenimiento de calificaciones máximas y mejoramiento de las calificaciones, y recursos y dependencia de la ayuda por otra parte. Análogamente, cabría esperar asociaciones positivas con comercio, PIB per cápita, crecimiento del PIB y democracia. Un análisis de correlaciones no arrojó conclusiones significativas. Como ya se examinó en el capítulo 2, se concluyó que no correspondía llevar a cabo regresiones multivariantes en esta etapa, dada la escasa magnitud de la muestra, la gama limitada de la variable dependiente y el carácter subjetivo de posibles variables explicativas. No obstante, sería útil, en una etapa ulterior, llevar a cabo una prueba en que se controle la posibilidad de que los países desarrollados no den muestras de una mayor mejora en la EC debido a que sus calificaciones ya eran altas en la EA, mediante la inclusión de calificaciones en la EA como variable explicativa en la regresión. También podrían analizarse factores que diferencien a los 32 países con EC de la muestra de países sin EC, que es mucho mayor. Las variables explicativas podrían incluir el volumen de la ayuda (o, en forma más ajustada, el respaldo presupuestario), la proporción de la ayuda que proviene del mayor de los donantes, la proporción de la ayuda que proviene de organismos asociados del PEFA, las dimensiones del país, la democracia e indicadores de la evaluación de las políticas e instituciones nacionales, etc.

100. La falta de conclusiones significativas puede obedecer a que las características del país pueden estar sujetas a dos influencias contrapuestas: por un lado mayor riqueza, mayor capacidad, régimen democrático, acelerado crecimiento, apertura al comercio, son factores de los que cabría esperar que se avanzara hacia el mejoramiento de la GFP, pero, por otro lado, esos países pueden ya haber alcanzado, por esas razones, un alto nivel en la EA, por lo cual podría ser difícil que siguieran progresando. Análogamente, los países con recursos naturales y dependientes de la ayuda podrían tener menos incentivos para seguir mejorando la GFP pero, por otro lado, por esas razones pueden haber partido de un nivel tan bajo en la EA que estén en condiciones de demostrar progreso.

101. En conclusión, los atributos formales de la GFP en que puede lograrse progreso a través de la adopción de nuevas leyes, reglamentos o instrumentos técnicos, o centrando la atención en tan solo unos pocos organismos o en una etapa temprana del ciclo presupuestario, tienen mayor probabilidad de mejorar o mantener un calificación máxima que los atributos funcionales de la GFP, en que el progreso requiere la efectiva aplicación de una nueva ley o reglamento o la coordinación de la labor de muchos organismos, o una labor realizada en etapas ulteriores del ciclo presupuestario. La diferencia es sumamente pronunciada en el caso de los atributos de la GFP en que puede lograrse progreso trabajando con uno o con unos pocos organismos, en comparación con atributos de la GFP en que participan muchos organismos. Análogamente, la probabilidad de que los atributos funcionales se deterioren o permanezcan en un nivel mínimo es mucho mayor en el caso de los atributos funcionales que en el de los formales. Uno u otro tipo de atributos tienen mayores proporciones de calificaciones máximas y crecientes que de calificaciones mínimas y en proceso de deterioro, aunque las diferencias entre los atributos formales son mayores que las que existen entre los funcionales. Estos resultados de las modalidades dinámicas de las calificaciones del PEFA en general están en consonancia con los resultados estáticos. No obstante, algunos aspectos de los resultados, tales como la mayor proporción de calificaciones crecientes o máximas en comparación con calificaciones en proceso de deterioro o mínimas para los atributos funcionales, son alentadores, pues demuestran que aun en las esferas de mayores dificultades es posible avanzar.

102. Se realizó un análisis ulterior sobre casos de intervalos breves, que demostró que en contextos propicios es posible lograr rápidamente ciertos tipos de mejoras. Finalmente, no pudieron encontrarse modalidades estadísticamente significativas de cambio entre diferentes tipos de características de países.

103. Determinar las razones que explican las tendencias aquí identificadas —es decir si las tendencias en la GFP obedecieron principalmente a la capacidad temporal proporcionada por expertos externos, a modificaciones del entorno macroeconómico o a genuinas reformas gubernamentales— rebasa el horizonte del presente informe, pero la cuestión puede examinarse en relación con el estudio de evaluación de donantes en curso sobre el respaldo para la reforma de la GFP. Para ello podría ser útil acudir al acervo de datos de ER del PEFA, que aumenta rápidamente.

Capítulo 5: Recomendaciones

104. Sobre la base de esas conclusiones se formularon recomendaciones, varias de las cuales se presentaron en anteriores informes de seguimiento. *La Secretaría del PEFA* debería adoptar las siguientes medidas:

- Revisar los documentos referentes al proceso de evaluación (“Lista de comprobación de términos de referencia”, “Buenas prácticas en la aplicación del marco de referencia para la medición del desempeño en materia de GFP”, “Nota de orientación para las evaluaciones periódicas”), para destacar la necesidad de: i) incluir en las notas conceptuales y los términos de referencia una referencia específica a las notas de orientación de la secretaría; ii) planificar la obtención de tiempo y recursos adicionales para analizar cambios; iii) presentar las notas conceptuales y los términos de referencia a la secretaría para obtener sus comentarios, y iv) transferir información detallada de la EA a los evaluadores de la EC mediante cierta coincidencia entre miembros de equipos, o por lo menos mediante la colaboración del jefe del equipo de la EA y comentarios de los interesados y de la Secretaría del PEFA sobre la EA.
- Revisar el material de capacitación para destacar los temas planteados en el punto anterior.
- Examinar las dimensiones de los indicadores con grandes dificultades de comparabilidad para establecer si es necesario describir más claramente el marco y brindar orientación a los evaluadores y determinar si debería considerarse la posibilidad de modificar los requisitos mínimos para la calificación de las dimensiones.

105. Los *organismos principales* deberían hacer cumplir ciertas buenas prácticas:

- En las notas conceptuales y los términos de referencia debería establecerse lo más específicamente posible lo que se espera de la ER, tal como se recomienda en la “Nota de orientación para las evaluaciones reiteradas” de la Secretaría del PEFA.
- En las notas conceptuales y los términos de referencia debería especificarse en forma suficientemente detallada la manera en que los evaluadores deberían incluir comparaciones con la EA en el informe de la EC y conceder tiempo para verificar las razones que hayan llevado a adjudicar anteriores calificaciones.
- En las notas conceptuales y los términos de referencia debería disponerse que el equipo de evaluación registre toda la evaluación pertinente (por ejemplo en un CD) en forma comprensible y de fácil acceso para otros expertos.
- Las notas conceptuales y los términos de referencia deberían incluirse como anexo del borrador o del texto final de los informes de evaluación.
- Las notas conceptuales y los términos de referencia deberían ser sometidos a un proceso de garantía de calidad (examen por los pares) a cargo de la Secretaría del PEFA, como garantía de adecuada implementación de las recomendaciones que anteceden.

- Debería seguirse la práctica de proporcionar al equipo de evaluadores de la EC toda la información y los documentos pertinentes surgidos de las evaluaciones anteriores, incluidos el informe final y comentarios de los interesados y de la secretaría.
- Los organismos principales deberían facilitar el contacto entre los jefes de equipos de evaluación anteriores y actuales, aunque ello requiera más tiempo e implique costos adicionales.
- Deberían proporcionarse a la secretaría razones específicas por las que no se compartan sus comentarios a las notas conceptuales y los términos de referencia, así como a las evaluaciones.

106. *Al realizar una evaluación reiterada, los evaluadores deberían tener en cuenta las siguientes recomendaciones:*

- Observar las pautas contenidas en la “Nota de orientación sobre evaluaciones reiteradas” de la Secretaría del PEFA.
- Solicitar a la entidad encargada de la gestión de la evaluación (el organismo principal) información sobre la EA (borradores de informes e informes finales, y comentarios del proceso de examen de calidad).
- Solicitar al organismo principal que se ponga en contacto con el equipo de EA.

Anexo A: Evaluaciones comparativas del gasto público y rendición de cuentas financieras

Región	País	Primera evaluación		Primera evaluación reiterada		Segunda evaluación reiterada	
		Organismo principal	Fecha del informe	Organismo principal	Fecha del informe	Organismo principal	Fecha del informe
África	Burkina Faso	Comisión Europea	Abr. 07	Gobierno	Jun. 10		
África	Etiopía	Comisión Europea	Oct. 07	Comisión Europea	Oct. 10		
África	Etiopía-región de Benishangul	Comisión Europea	Oct. 07	Comisión Europea	Jul. 10		
África	Etiopía-región de Harari	Comisión Europea	Oct. 07	Comisión Europea	Jul. 10		
África	Etiopía-región de Oromiya	Comisión Europea	Oct. 07	Comisión Europea	Jul. 10		
África	Ghana	Banco Mundial	Jun. 06	Comisión Europea	Comisión Europea		
África	Guinea Bissau	Banco Mundial	Jun. 06	Comisión Europea	Mayo 09		
África	Kenya	DFID	Jul. 06	Comisión Europea	Mar. 09		
África	Lesotho	Banco Mundial	Jun. 07	DFID	Jul. 09		
África	Madagascar	Comisión Europea	Mayo 06	Banco Mundial	Mayo 08		
África	Malawi	Comisión Europea	Jul. 05	Comisión Europea	Ago. 06	Comisión Europea	Jun. 08
África	Mozambique	Comisión Europea	Mar. 06	NORAD	Feb. 08		
África	Sierra Leona	DFID	Dic. 07	DFID	Sep.10		
África	Swazilandia	Comisión Europea	Ene. 07	Banco Mundial	Mayo 10		
África	Tanzanía	Banco Mundial	Mayo 06	DFID	Mayo 10		
África	Uganda	Comisión Europea	Mayo 06	Banco Mundial	Jun. 09		
África	Zambia	DFID	Dic. 05	Gobierno	Jun. 08		
Asia oriental y el Pacífico	Samoa	Comisión Europea	Oct. 06	Gobierno	Abr.10		
Asia oriental y el Pacífico	Timor Leste	Comisión Europea	Feb. 07	FMI	Jun. 10		

Asia oriental y el Pacífico	Tonga	OADI	Sep. 07	OADI	Mayo 10		
Asia oriental y el Pacífico	Vanuatu	Comisión Europea	Jul. 06	Comisión Europea	Nov. 09		
Europa y Asia central	Kosovo	Banco Mundial	Mar. 07	Gobierno	Mayo 09		
Europa y Asia central	República Kirguisa	DFID	Ene. 06	SECO	Dic. 09		
Europa y Asia central	Moldova	Comisión Europea	Abr. 06	Banco Mundial	Jul. 08		
Europa y Asia central	Serbia	Banco Mundial	Feb. 07	Gobierno	Sep. 10		
América Latina y el Caribe	Barbados	Comisión Europea	Oct. 06	Comisión Europea	Jul. 10		
América Latina y el Caribe	Dominica	Comisión Europea	Abr. 07	Comisión Europea	Jun. 10		
América Latina y el Caribe	República Dominicana	Comisión Europea	Nov. 07	Comisión Europea	Sep.10		
América Latina y el Caribe	Saint Kitts y Nevis	Comisión Europea	Abr. 07	Comisión Europea	Dic. 09		
América Latina y el Caribe	Santa Lucía	Comisión Europea	Oct. 06	Comisión Europea	Feb.10		
América Latina y el Caribe	Trinidad y Tabago	Comisión Europea	Jun. 06	Comisión Europea	Dic. 08		
Asia meridional	Afganistán	Banco Mundial	Dic. 05	Banco Mundial	Jun. 08		

Anexo B: Evaluaciones reiteradas del gasto público y rendición de cuentas financieras no consideradas “comparativas”

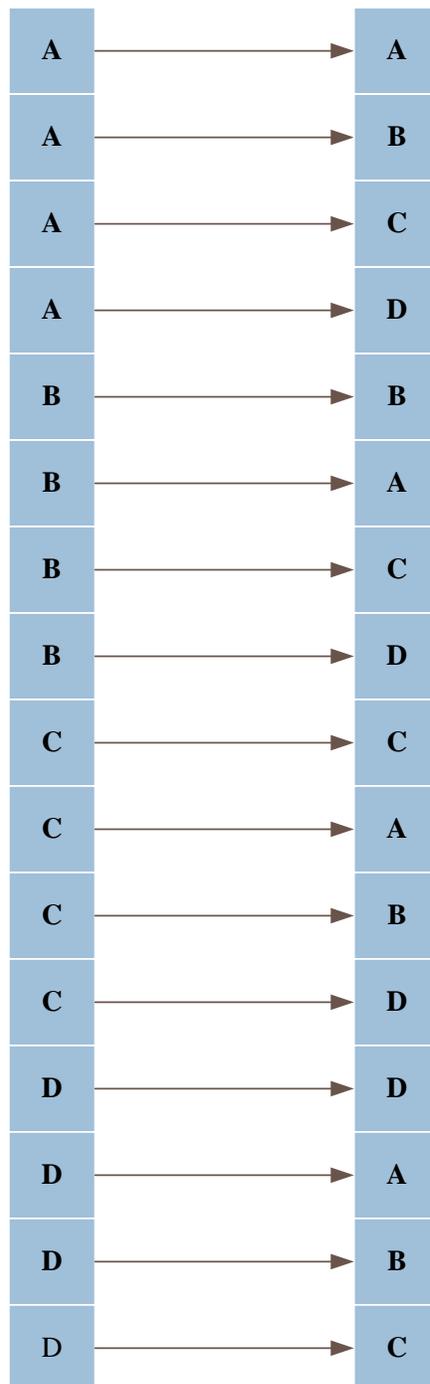
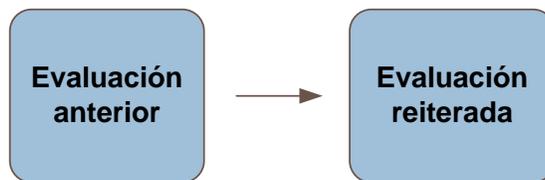
Región	País	Primera evaluación		Segunda evaluación	
		Organismo principal	Fecha del informe	Organismo principal	Fecha del informe
África	República Centroafricana	Banco Mundial	Jun. 08	Comisión Europea	Jul. 10
África	Ghana	DFID	Sep. 05	Banco Mundial	Jun. 06
África	Santo Tomé y Príncipe	Banco Mundial	Jun. 07	Comisión Europea	Ene. 10
África	Togo	Banco Mundial	Jun. 06	Comisión Europea	Mar. 09
África	Uganda	Comisión Europea	Mayo 06	Auditor general	Mar. 08*
Asia oriental y el Pacífico	República Democrática Popular Lao	Comisión Europea	Mayo 06	Banco Mundial	Jun. 10
Asia oriental y el Pacífico	Papua Nueva Guinea	Banco Mundial	Sep. 05	Banco Mundial	Mar. 09
América Latina y el Caribe	Bolivia	Gobierno	Oct. 07*	Banco Mundial	Ago. 09
América Latina y el Caribe	República Dominicana	Comisión Europea	Nov. 07	Gobierno	Nov. 09*
América Latina y el Caribe	Granada	Comisión Europea	Sep. 06	Comisión Europea	Mar. 10
América Latina y el Caribe	Honduras	Banco Mundial	Jun. 06	Comisión Europea	Abr. 09
Otras	Reino Unido-Montserrat	Comisión Europea	Sep. 08*	FMI	Dic. 09

* La secretaría no recibió de interesados en el proceso de evaluación solicitudes de realización de un examen del borrador del informe de evaluación del PEFA, por lo cual no lo realizó ni proporcionó comentarios al organismo principal.

- República Centroafricana: la ER no mide los cambios en el desempeño. No se justifica la modificación de la calificación efectuada en el cuadro.
- Ghana: en el informe de junio de 2006 no se hizo referencia al informe de septiembre de 2005.
- Santo Tomé y Príncipe: la ER no se refiere a la EA de 2006.
- Togo: la EA no fue un PEFA (exámenes del gasto público/evaluación de la capacidad de gestión financiera/PEFA sectoriales). La ER no mide los cambios en el desempeño.
- Uganda: en la evaluación no se utilizó la evaluación de 2005 para medir los cambios en el desempeño.

- Lao: no se citaron ni utilizaron las calificaciones de la evaluación parcial del PEFA de 2006 en la evaluación de 2010.
- Papua Nueva Guinea: el informe no contiene comparaciones con calificaciones anteriores del PEFA.
- Bolivia: en la ER no se utilizó la “autoevaluación” de 2007 para medir los cambios en el desempeño; se menciona la autoevaluación de 2007 pero no se prevé una comparación.
- República Dominicana: se trata de una autoevaluación presentada como entrenamiento para autoridades públicas antes de la ER de 2010. No se hace referencia a la EA, de 2007.
- Granada: en la ER no se hizo referencia a la EA.
- Honduras: en el PEFA de 2009 no se midieron los cambios en el desempeño a lo largo del tiempo.
- Reino Unido-Montserrat: en 2008 se realizó una evaluación del PEFA, pero no se hizo referencia a las calificaciones en la evaluación de 2009.

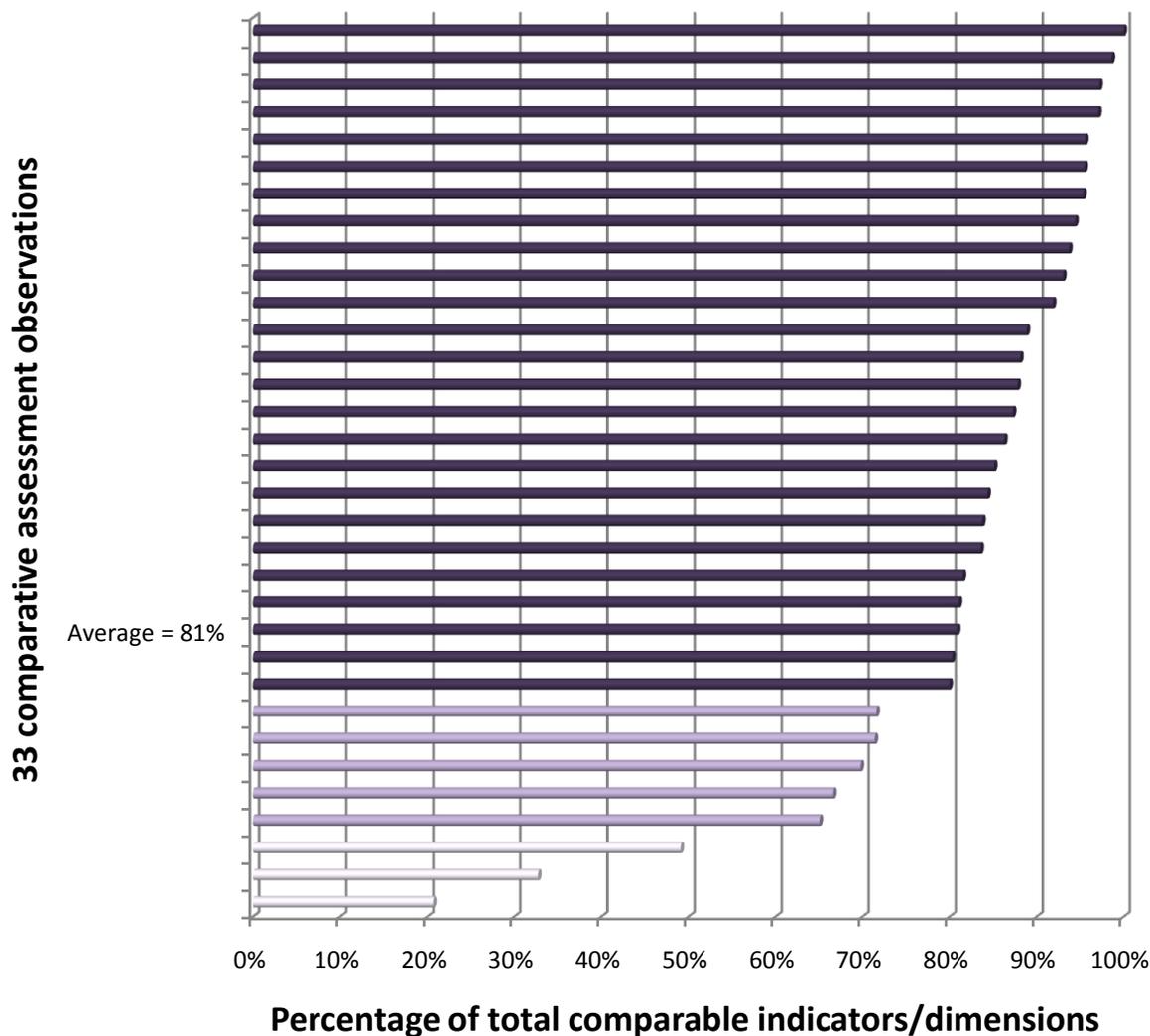
Anexo C: Combinaciones de calificaciones de dimensiones entre una evaluación anterior y una evaluación reiterada



Anexo D: Países con evaluación inicial completada antes del 30 de junio de 2006 (es decir hace más de cuatro años y medio) que aún no han completado una evaluación reiterada

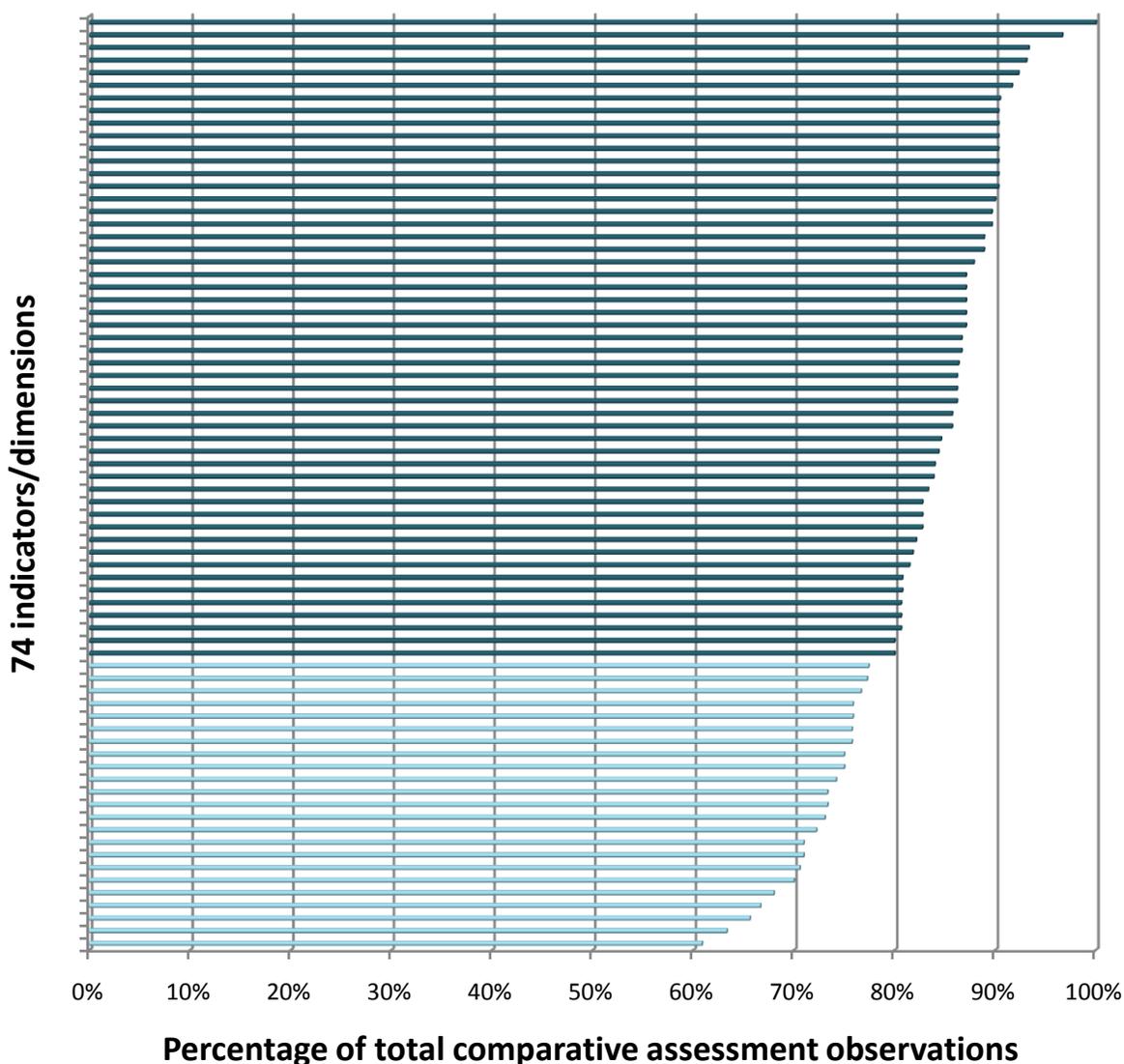
País	Donante principal	Publicada	Estado	Fecha indicada en la página inicial del informe	Estado actual de la evaluación comparativa
Bangladesh	Banco Mundial	Sí	Finalizada	Abr. 06	En curso
República del Congo	Comisión Europea	Sí	Finalizada	Mar. 06	No se dispone de información
Siria	FMI	No	Finalizada	Mar. 06	No se dispone de información
Uganda-Gobierno local	Comisión Europea	No	Sustancialmente finalizada	Dic. 05	Planificada
Islas Fiji	Banco Mundial	No	Finalizada	Jun. 05	No se dispone de información

Anexo E: Porcentaje de dimensiones de indicadores que pueden compararse con confianza entre distintas evaluaciones comparativas, por evaluación, si se prescinde de los casos “sin calificar”



33 comparative assessment observations	Observaciones sobre 33 evaluaciones comparativas
Average	Promedio
Percentage of total comparable indicators/dimensions	Porcentaje del total de indicadores/dimensiones comparables

Anexo F: Porcentaje de calificaciones que pueden compararse con confianza entre distintas evaluaciones comparativas para cada dimensión del indicador si se prescinde de los casos “sin calificar”



Leyenda³¹:

100%-	D-1 i), ID-14 ii), ID-26 iii), ID-14 iii), ID-12 ii), D-1 ii), D-3, ID-11 i), ID-11 iii), ID-13 ii), ID-16 ii), ID-28 i), ID-28 ii), ID-28 iii), ID-25 ii), ID-18 i), ID-21 ii), ID-17 i), ID-17 iii), ID-10, ID-11 ii), ID-12 iii), ID-22 ii), ID-24 ii), ID-24 iii), ID-14 i), ID-23, D-2 ii), ID-15 ii), ID-21 i), ID-21 iii), ID-27 iii), D-2 i), ID-8 iii), ID-22 i), ID-8 ii), ID-16 i), ID-26 i), ID-17 ii), ID-18 ii), ID-18 iii), ID-27 v), ID-6, ID-15 iii), ID-4 ii), ID-9 i), ID-12 i), ID-24 i), ID-25 i), ID-13 iii), ID-20 iii)
80%	

³¹ Los indicadores se presentan en orden descendente. El indicador D-1 i), por ejemplo, es el 100%, en tanto que ID-20 iii) es el más próximo al 80%.

79%- 60%	ID-12 iv), ID-9 ii), ID-25 iii), ID-18 iv), ID-19 i), ID-3, ID-5, ID-27 i), ID-27 ii), ID-13 i), ID-19 iii), ID-20 i), ID-7 ii), ID-15 i), ID-2, ID-16 iii), ID-4 i), ID-26 ii), ID-8 i), ID-20 ii), ID-1, ID-19 ii), ID-7 i)
-------------	---

74 indicators/dimensions	74 indicadores/dimensiones
Percentage of total comparative assessment observations	Porcentaje del total de observaciones de evaluaciones comparativas

Anexo G: Clasificación porcentual de evaluaciones comparativas comparables, sin calificación y no comparables

Indicador/ dimensión	Comparada con confianza	Sin califica r	No comparable
ID-1	64%	3%	33%
ID-2	67%	6%	27%
ID-3	76%	0%	24%
ID-4 i)	36%	48%	15%
ID-4 ii)	64%	21%	15%
ID-5	76%	0%	24%
ID-6	82%	0%	18%
ID-7 i)	42%	30%	27%
ID-7 ii)	58%	21%	21%
ID-8 i)	52%	24%	24%
ID-8 ii)	64%	24%	12%
ID-8 iii)	67%	21%	12%
ID-9 i)	64%	21%	15%
ID-9 ii)	52%	33%	15%
ID-10	88%	0%	12%
ID-11 i)	85%	6%	9%
ID-11 ii)	82%	6%	12%
ID-11 iii)	85%	6%	9%
ID-12 i)	76%	6%	18%
ID-12 ii)	73%	21%	6%
ID-12 iii)	82%	6%	12%
ID-12 iv)	73%	6%	21%
ID-13 i)	70%	6%	24%
ID-13 ii)	85%	6%	9%
ID-13 iii)	73%	9%	18%
ID-14 i)	79%	9%	12%
ID-14 ii)	88%	9%	3%
ID-14 iii)	82%	12%	6%
ID-15 i)	39%	45%	15%
ID-15 ii)	76%	12%	12%
ID-15 iii)	67%	18%	15%
ID-16 i)	79%	6%	15%
ID-16 ii)	85%	6%	9%
ID-16 iii)	67%	6%	27%
ID-17 i)	73%	18%	9%
ID-17 ii)	73%	12%	15%
ID-17 iii)	73%	18%	9%

Indicador/ dimensión	Comparada con confianza	Sin calific ar	No comparable
ID-18 i)	79%	12%	9%
ID-18 ii)	73%	12%	15%
ID-18 iii)	73%	12%	15%
ID-18 iv)	67%	12%	21%
ID-19 i)	67%	12%	21%
ID-19 ii)	58%	9%	33%
ID-19 iii)	67%	9%	24%
ID-20 i)	67%	9%	24%
ID-20 ii)	61%	9%	30%
ID-20 iii)	73%	9%	18%
ID-21 i)	76%	12%	12%
ID-21 ii)	79%	12%	9%
ID-21 iii)	76%	12%	12%
ID-22 i)	82%	3%	15%
ID-22 ii)	82%	6%	12%
ID-23	79%	9%	12%
ID-24 i)	76%	6%	18%
ID-24 ii)	82%	6%	12%
ID-24 iii)	82%	6%	12%
ID-25 i)	76%	6%	18%
ID-25 ii)	82%	9%	9%
ID-25 iii)	70%	9%	21%
ID-26 i)	76%	9%	15%
ID-26 ii)	64%	9%	27%
ID-26 iii)	85%	9%	6%
ID-27 i)	64%	15%	21%
ID-27 ii)	64%	15%	21%
ID-27 iii)	73%	15%	12%
ID-27 v)	70%	15%	15%
ID-28 i)	85%	6%	9%
ID-28 ii)	85%	6%	9%
ID-28 iii)	85%	6%	9%
D-1 i)	42%	58%	0%
D-1 ii)	33%	64%	3%
D-2 i)	55%	36%	9%
D-2 ii)	58%	33%	9%
D-3	58%	36%	6%

Anexo H: Clasificación porcentual de cambios en las evaluaciones comparativas en los distintos indicadores y dimensiones

Indicador/dimensión	"A" a "A"	Calificaciones crecientes	"D" a "D"	Calificaciones decrecientes	Mantenida en "B" o "C"	Sin calificar/no comparable
ID-1	18%	18%	6%	18%	3%	36%
ID-2	12%	12%	15%	9%	18%	33%
ID-3	45%	9%	0%	18%	3%	24%
ID-4 i)	21%	6%	3%	3%	3%	64%
ID-4 ii)	6%	21%	6%	12%	18%	36%
ID-5	9%	15%	0%	6%	45%	24%
ID-6	15%	36%	0%	0%	30%	18%
ID-7 i)	24%	3%	3%	9%	3%	58%
ID-7 ii)	9%	18%	6%	6%	18%	42%
ID-8 i)	27%	12%	3%	6%	3%	48%
ID-8 ii)	18%	15%	6%	15%	9%	36%
ID-8 iii)	12%	6%	39%	6%	3%	33%
ID-9 i)	0%	12%	3%	9%	39%	36%
ID-9 ii)	15%	12%	18%	0%	6%	48%
ID-10	3%	27%	3%	9%	45%	12%
ID-11 i)	21%	9%	0%	15%	39%	15%
ID-11 ii)	21%	27%	0%	12%	21%	18%
ID-11 iii)	24%	21%	12%	12%	15%	15%
ID-12 i)	0%	24%	0%	6%	45%	24%
ID-12 ii)	18%	39%	3%	6%	6%	27%
ID-12 iii)	0%	27%	24%	15%	15%	18%
ID-12 iv)	0%	24%	18%	9%	21%	27%
ID-13 i)	9%	12%	0%	6%	42%	30%
ID-13 ii)	15%	48%	0%	3%	18%	15%
ID-13 iii)	6%	24%	0%	3%	39%	27%
ID-14 i)	3%	21%	3%	3%	48%	21%
ID-14 ii)	3%	39%	0%	15%	30%	12%

Indicador/dimensión	"A" a "A"	Calificaciones crecientes	"D" a "D"	Calificaciones decrecientes	Mantenida en "B" o "C"	Sin calificar/no comparable
ID-18 i)	21%	18%	15%	15%	9%	21%
ID-18 ii)	15%	24%	6%	9%	18%	27%
ID-18 iii)	21%	27%	3%	6%	15%	27%
ID-18 iv)	0%	21%	6%	12%	27%	33%
ID-19 i)	0%	30%	27%	6%	3%	33%
ID-19 ii)	0%	15%	3%	12%	27%	42%
ID-19 iii)	3%	18%	15%	9%	21%	33%
ID-20 i)	3%	24%	6%	9%	24%	33%
ID-20 ii)	0%	18%	6%	3%	33%	39%
ID-20 iii)	0%	21%	3%	6%	42%	27%
ID-21 i)	6%	21%	27%	0%	21%	24%
ID-21 ii)	0%	33%	12%	12%	21%	21%
ID-21 iii)	3%	18%	27%	12%	15%	24%
ID-22 i)	6%	33%	6%	6%	30%	18%
ID-22 ii)	6%	30%	15%	6%	24%	18%
ID-23	6%	9%	33%	21%	9%	21%
ID-24 i)	21%	18%	0%	12%	24%	24%
ID-24 ii)	48%	21%	0%	3%	9%	18%
ID-24 iii)	21%	24%	3%	9%	24%	18%
ID-25 i)	12%	18%	9%	15%	21%	24%
ID-25 ii)	30%	27%	9%	6%	9%	18%
ID-25 iii)	12%	21%	6%	3%	27%	30%
ID-26 i)	0%	30%	12%	9%	24%	24%
ID-26 ii)	0%	30%	9%	12%	12%	36%
ID-26 iii)	3%	24%	15%	9%	33%	15%
ID-27 i)	9%	15%	0%	6%	33%	36%
ID-27 ii)	6%	24%	6%	6%	21%	36%

ID-14 iii)	0%	42%	3%	6%	30%	18%
ID-15 i)	3%	9%	21%	6%	0%	61%
ID-15 ii)	36%	21%	0%	3%	15%	24%
ID-15 iii)	18%	18%	18%	9%	3%	33%
ID-16 i)	15%	30%	6%	6%	21%	21%
ID-16 ii)	3%	27%	6%	15%	33%	15%
ID-16 iii)	12%	9%	3%	6%	36%	33%
ID-17 i)	9%	21%	0%	3%	39%	27%
ID-17 ii)	6%	30%	0%	18%	18%	27%
ID-17 iii)	12%	24%	0%	6%	30%	27%

ID-27 iii)	18%	15%	15%	9%	15%	27%
ID-27 v)	3%	6%	0%	18%	42%	30%
ID-28 i)	15%	12%	39%	15%	3%	15%
ID-28 ii)	6%	24%	30%	12%	12%	15%
ID-28 iii)	0%	12%	33%	6%	33%	15%
D-1 i)	15%	12%	6%	9%	0%	58%
D-1 ii)	3%	12%	9%	9%	0%	67%
D-2 i)	3%	12%	15%	6%	18%	45%
D-2 ii)	0%	18%	21%	3%	15%	42%
D-3	0%	9%	39%	6%	3%	42%

Anexo I: Metodología de codificación de la gestión de las finanzas públicas elaborada por M. Andrews

Número de dimensión del PEFA	<i>De jure</i>	<i>De facto</i>	Concentrada	Desconcentrada	Inicial	Ulterior
ID-5	1	0	1	0	1	0
ID-6	1	0	1	0	1	0
ID-7 i)	0	1	0	1	0	1
ID-7 ii)	0	1	0	1	0	1
ID-8 i)	1	0	1	0	1	0
ID-8 ii)	0	1	0	1	1	0
ID-8 iii)	0	1	0	1	0	1
ID-9 i)	0	1	0	1	0	1
ID-9 ii)	0	1	0	1	0	1
ID-10	1	0	1	0	1	0
ID-11 i)	1	0	1	0	1	0
ID-11 ii)	1	0	1	0	1	0
ID-11 iii)	0	1	0	1	1	0
ID-12 i)	0	1	1	0	1	0
ID-12 ii)	1	0	1	0	1	0
ID-12 iii)	0	1	0	1	1	0
ID-12 iv)	0	1	0	1	1	0
ID-13 i)	1	0	1	0	0	1
ID-13 ii)	1	0	1	0	0	1
ID-13 iii)	1	0	1	0	0	1
ID-14 i)	1	0	1	0	0	1
ID-14 ii)	0	1	1	0	0	1
ID-14 iii)	1	0	1	0	0	1
ID-15 i)	0	1	0	1	0	1
ID-15 ii)	0	1	1	0	0	1
ID-15 iii)	0	1	1	0	0	1
ID-16 i)	1	0	1	0	0	1
ID-16 ii)	0	1	0	1	0	1
ID-16 iii)	0	1	0	1	0	1
ID-17 i)	1	0	1	0	0	1
ID-17 ii)	1	0	1	0	0	1
ID-17 iii)	1	0	1	0	0	1
ID-18 i)	0	1	0	1	0	1
ID-18 ii)	0	1	0	1	0	1
ID-18 iii)	1	0	0	1	0	1

ID-18 iv)	0	1	0	1	0	1
ID-19 i)	0	1	0	1	0	1
ID-19 ii)	0	1	0	1	0	1
ID-19 iii)	1	0	0	1	0	1
ID-20 i)	0	1	0	1	0	1
ID-20 ii)	1	0	0	1	0	1
ID-20 iii)	0	1	0	1	0	1
ID-21 i)	0	1	0	1	0	1
ID-21 ii)	1	0	0	1	0	1
ID-21 iii)	0	1	0	1	0	1
ID-22 i)	0	1	1	0	0	1
ID-22 ii)	0	1	0	1	0	1
ID-23	0	1	0	1	0	1
ID-24 i)	1	0	1	0	0	1
ID-24 ii)	0	1	1	0	0	1
ID-24 iii)	0	1	1	0	0	1
ID-25 i)	0	1	0	1	0	1
ID-25 ii)	0	1	1	0	0	1
ID-25 iii)	1	0	1	0	0	1
ID-26 i)	0	1	0	1	0	1
ID-26 ii)	0	1	0	1	0	1
ID-26 ii)	0	1	0	1	0	1
ID-27 i)	1	0	0	1	1	0
ID-27 ii)	1	0	0	1	1	0
ID-27 iii)	1	0	0	1	1	0
ID-27 iv)	1	0	0	1	1	0
ID-28 i)	0	1	0	1	0	1
ID-28 ii)	0	1	0	1	0	1
ID-28 iii)	0	1	0	1	0	1
Totales	26	38	26	38	16	48

Anexo J: Referencias

Andrews, Matt, 2009. Isomorphism and the Limits to African Public Financial Management Reform, RWP09-012, Kennedy School, Harvard. Fecha de acceso: 11 de mayo de 2011, <http://web.hks.harvard.edu/publications/getFile.aspx?Id=340>.

De Renzio, Paulo, 2009. Taking Stock: What do PEFA Assessments tell us about PFM systems across countries? Working Paper 302. Londres, Overseas Development Institute. Fecha de acceso: 11 de mayo de 2011 <http://siteresources.worldbank.org/PEFA/Resources/TakingStockRenzio2007.pdf>.

Departamento de Desarrollo Internacional del Reino Unido, 2009. Fiduciary Risk Assessment. Londres: DFID.

Gobierno de la República de Bangladesh y Departamento de Desarrollo Internacional del Reino Unido, 2007. Assessment of the Impact of Financial Management Reforms in Bangladesh 1992 to 2006.

Mackie, Andrew y Caprio, Giovanni, 2010. Assessing the Impact of the PEFA Framework, estudio para el borrador del informe de SC del PEFA, noviembre.

Porter, Doug, Matt Andrews, Joel Turkewitz y Clay Wescott, 2010. Managing Public Finance and Procurement in Fragile and Conflicted Settings. Washington, DC, Banco Mundial. Fecha de acceso: 11 de mayo de 2011, <http://wdr2011.worldbank.org/procurement>.

Banco Mundial, 2006. Bangladesh Country Assistance Strategy 2006-2009. Washington, DC, Banco Mundial. Fecha de acceso: 11 de mayo de 2011, http://siteresources.worldbank.org/BANGLADESHEXTN/Resources/CAS_MAIN_BOOK_FINAL.pdf.

Banco Mundial, 2007. Public Sector Accounting and Auditing Gap Analysis. Washington, DC, Banco Mundial

Banco Mundial, 2009. Bangladesh Public Expenditure and Institutional Review. Washington, DC, Banco Mundial. Fecha de acceso: 11 de mayo de 2011 [http://www.spemp.com/admin/download.php?f=Public%20Expenditure%20and%20Institutional%20Review%20\(PEIR%20-%20I\),%202010.pdf](http://www.spemp.com/admin/download.php?f=Public%20Expenditure%20and%20Institutional%20Review%20(PEIR%20-%20I),%202010.pdf).