



**Cuestiones relativas a la comparación y la agregación
de los resultados de las evaluaciones del PEFA
en distintos países a lo largo del tiempo**

Versión final
Secretaría del PEFA
13 de mayo de 2009

ÍNDICE

Página n.º

Resumen	3
1. Introducción.....	5
2. Finalidad de las comparaciones en los países y entre países.....	6
3. Ejemplos de métodos utilizados en las comparaciones y las agregaciones.....	7
3.1 Comparación en los países	7
3.2 Comparación entre países	10
3.3 Comparación simultánea entre países a lo largo del tiempo	10
4. Cuestiones derivadas de la agregación y la comparación	13
4.1 Agregación de los distintos indicadores	13
4.2 Conversión de una escala ordinal a una escala numérica.....	15
4.3 Agrupación de países.....	17
4.4 Formas de reducir los problemas de comparabilidad	21
5. Conclusiones.....	23
APÉNDICE 1 Bibliografía.....	25

Resumen

- i. Las evaluaciones del Programa de Gasto Público y Rendición de Cuentas Financieras (PEFA) se llevan a cabo para identificar los puntos fuertes y débiles de los sistemas de gestión de las finanzas públicas (GFP) de un determinado país y establecer parámetros de desempeño aplicables a la formulación de programas de reforma y al seguimiento de los progresos en materia de desempeño a lo largo del tiempo.
- ii. Dado que actualmente se dispone de datos completos sobre el desempeño de varios países en materia de GFP, cabe preguntarse cómo deben utilizarse esos datos para realizar comparaciones significativas de calificaciones en los países a lo largo del tiempo y comparaciones entre distintos países tanto en un momento determinado como a través del tiempo.
- iii. El enfoque más básico —y recomendado universalmente— para hacerlo consistiría en comparar las calificaciones de cada indicador y utilizar la parte descriptiva de los informes de desempeño en materia de GFP a fin de extraer conclusiones sobre los motivos de las diferencias en las calificaciones. Este tipo de comparación ayuda a interpretar estas diferencias a la luz de las circunstancias del país de que se trate. Si bien un análisis de esta naturaleza sería un requisito estándar para realizar una comparación del desempeño a lo largo del tiempo en un país, podría resultar complicado e insumir mucho tiempo en el caso del análisis de varios países. Sin embargo, el análisis no conduce en ninguno de los casos a medidas sencillas que sean fáciles de comprender para quienes no son especialistas y que puedan transmitirse fácilmente a los encargados de tomar las decisiones. Esto último reviste particular interés cuando se compara un número considerable de evaluaciones.
- iv. Por ello, autores de informes de evaluación, investigadores y otros usuarios han procurado encontrar una medida agregada del desempeño en materia de GFP para facilitar la comparación de las calificaciones entre los países y, en algunos casos, facilitar un mejor desempeño a lo largo del tiempo en un país determinado. Este tipo de comparaciones ha despertado gran interés debido a que la disponibilidad de datos sobre los indicadores del PEFA está aumentando rápidamente. Mediante este programa, se están intensificando los esfuerzos para poner esos datos a disposición de los posibles usuarios.
- v. La agregación de los resultados de la evaluación de un país comúnmente entraña tanto la conversión de las calificaciones ordinales de los indicadores a valores numéricos como la asignación de ponderaciones a los distintos indicadores. Cuando se agrupan los resultados de los países, la agregación también requiere la asignación de ponderaciones a cada país.
- vi. En el presente documento se proporcionan ejemplos de comparaciones realizadas por varios usuarios y se analizan las cuestiones metodológicas conexas. No existe un método científicamente correcto para determinar cómo deberían realizarse la agregación y la comparación en cada uno de esos tres niveles de supuestos. En consecuencia, el PEFA no apoya ni recomienda un enfoque en particular con respecto a la agregación.

- vii. El PEFA recomienda a los usuarios que —como parte de la difusión de los resultados obtenidos en las comparaciones— se expliquen claramente el método de agregación y los supuestos aplicados a cada caso, así como las razones de la elección. También sería aconsejable que los usuarios se sometieran a un análisis de sensibilidad para determinar el grado de solidez de sus conclusiones a la luz de supuestos de agregación alternativos.

1. Introducción

1. Dado que actualmente se dispone de datos completos sobre el desempeño de varios países en materia de GFP, cabe preguntarse cómo deben utilizarse esos datos para realizar comparaciones significativas de calificaciones en los países a lo largo del tiempo y comparaciones entre distintos países tanto en un momento determinado como a través del tiempo.
2. El enfoque más básico para hacerlo consistiría en comparar las calificaciones de cada indicador y utilizar la parte descriptiva de los informes del PEFA (Informes de desempeño en materia de GFP) para extraer conclusiones sobre los motivos de las diferencias en las calificaciones. Ello insume tiempo y no conduce a una medida sencilla que sea fácil de comprender para quienes no son especialistas y pueda transmitirse fácilmente a los encargados de tomar las decisiones.
3. Por lo tanto, resulta interesante tratar de encontrar una medida agregada de los resultados que facilite la comparación de las calificaciones entre países, ya sea una medida general del desempeño agregado de los países o para cada una de las seis dimensiones fundamentales del Marco del PEFA (en adelante, el “Marco”).
4. Las calificaciones de las evaluaciones del PEFA ya se han utilizado en varios documentos de investigación y en informes de evaluación para facilitar dichas comparaciones, ya sea calculando las calificaciones numéricas agregadas correspondientes a las seis dimensiones fundamentales del Marco y/o agregando las calificaciones para crear una medida del desempeño general en materia de GFP.
5. La finalidad de este documento es ilustrar los métodos aplicados para comparar el desempeño de los sistemas de GFP entre países, utilizando calificaciones de los indicadores extraídas de los informes de evaluación del PEFA, y señalar los inconvenientes que podría generar la aplicación de esos métodos. El documento se centra en la comparación del desempeño de los sistemas de GFP de distintos países según la medición de las seis dimensiones fundamentales del Marco del PEFA¹. Sin embargo, las cuestiones planteadas también se aplican a los usuarios de los informes de evaluación del PEFA que quieran utilizar subgrupos distintos de indicadores o indicadores individuales.
6. La estructura del resto de esta nota es la siguiente:
 - Sección 2: una descripción de la finalidad de las comparaciones en los países y entre países.
 - Sección 3: algunos ejemplos de los métodos utilizados para realizar comparaciones y agregaciones.
 - Sección 4: un análisis de las cuestiones derivadas de la agregación y la comparación.
 - Sección 5: las conclusiones.

¹ Los indicadores son: Credibilidad del presupuesto (ID 1-4), amplitud y transparencia (ID5-10), presupuestación basada en políticas (ID 11-12), previsibilidad y control en la ejecución del presupuesto (ID 13-21), contabilidad, registro y presentación de informes (ID 22-25), y verificación y auditoría externas (ID 26-28).

2. Finalidad de las comparaciones en los países y entre países

7. Si bien el propósito fundamental del Marco del PEFA era seguir de cerca los avances logrados a lo largo del tiempo en un país determinado, el uso de dicho marco para comparar el desempeño de distintos países en materia de GFP satisface los intereses de varios agentes. Por lo tanto, los interesados se muestran cada vez más dispuestos a procurar una forma razonablemente manejable de realizar estas comparaciones (es decir, que no se reduzca a comparar las calificaciones de todos y cada uno de los indicadores del PEFA). Los intereses en juego son los siguientes:

- Los Gobiernos pueden estar interesados en hacer el seguimiento de su propio desempeño en materia de GFP a través del tiempo en función de las dimensiones fundamentales del Marco del PEFA y compararlo con el de los países vecinos o con países que se encuentran en etapas similares o diferentes de la reforma de la GFP y/o de desarrollo económico en general.
- Los organismos donantes pueden estar interesados en hacer el seguimiento de su propio desempeño en materia de GFP a través del tiempo en sus países asociados y compararlo con el de otros países asociados a fin de seguir de cerca los avances logrados en sus operaciones de ayuda o extraer enseñanzas que resulten útiles a la hora de brindar asesoramiento y asistencia a los Gobiernos en la elaboración/revisión de sus programas de reforma de la GFP.
- La agrupación de las medidas del desempeño en materia de GFP de distintos países es también un requisito establecido en la Declaración de París sobre la Eficacia de la Ayuda al Desarrollo (indicadores 2a, 5 y 7) en relación con el seguimiento mundial de los progresos en materia de GFP.
- Por último, los académicos pueden estar interesados en explicar las varianzas de desempeño en materia de GFP estableciendo correlaciones entre las diferencias en el desempeño que muestran distintos países en ese ámbito y las variables explicativas que posiblemente reflejen las características de los países, por ejemplo los ingresos per cápita y el tamaño de la población.

3. Ejemplos de métodos utilizados en las comparaciones y las agregaciones

3.1 Comparación en los países

8. El Marco del PEFA se creó con el objetivo principal de medir los progresos a lo largo del tiempo. Durante 2008, comenzaron a realizarse evaluaciones reiteradas para hacer el seguimiento de dichos progresos². A fin de transmitir mejor la información sobre los cambios en el desempeño registrados a lo largo del tiempo, los autores de los informes han tratado de simplificar los resultados en menos medidas que la enumeración de las calificaciones correspondientes a los 28+3 indicadores en dos años, junto con un análisis descriptivo de la comparación. Se han observado dos métodos de agregación.
9. Uno de los métodos ha sido computar el número y la magnitud de los cambios en las calificaciones y sumarlos. En el cuadro 1 se proporciona un ejemplo simplificado. La medida resultante es muy sencilla y fácil de entender para quienes no son especialistas, pero se basa en supuestos relacionados con la conversión de la escala ordinal a intervalos numéricos y con la asignación de ponderaciones a los indicadores. En el ejemplo, ¿la mejora del ID-11 tiene un valor igual o mayor que el empeoramiento que se observa en el ID-14? ¿El cambio de la categoría “sin calificación” a D+ debería considerarse o no una mejora en el desempeño? Y, por lo tanto, ¿qué significa exactamente una mejora agregada de un punto en ocho indicadores?

Cuadro 1 Ejemplo ilustrativo de cuestiones derivadas de la agregación de los cambios en el desempeño

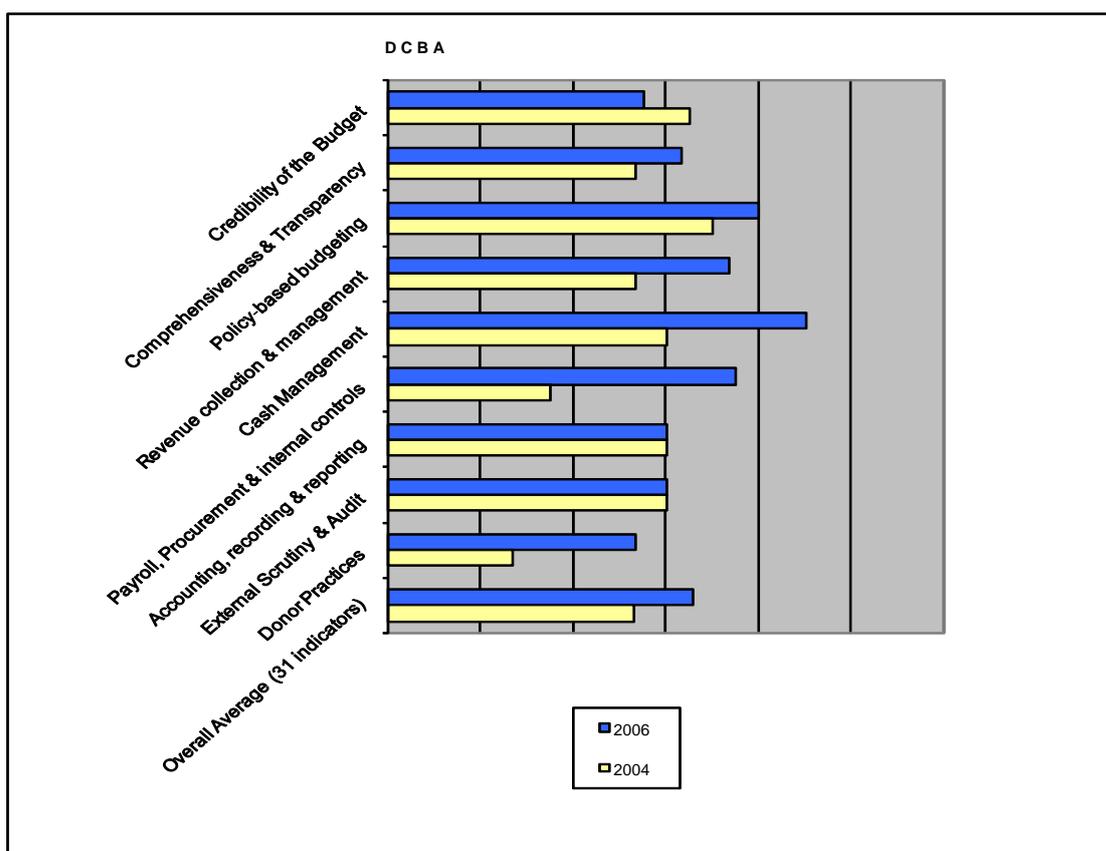
Indicador	Calificación de 2006	Calificación de 2009	Cambio
ID-11	C+	B	+ ½
ID-12	D	C	+ 1
ID-13	Sin calificación	D+	n.a.
ID-14	B+	B	- ½
ID-15	C	C+	+ ½
ID-16	C+	D+	- 1
ID-17	B	B	Cero
ID-18	D+	C	+ ½
Cambio general 1/			+ 1

1/ Se supone que existe un mismo intervalo numérico entre todas las calificaciones de la escala ordinal y una misma ponderación de indicadores.

² Existen algunas evaluaciones reiteradas anteriores, pero o bien se realizaron demasiado pronto luego del punto de referencia como para brindar una medición adecuada de los progresos en el tiempo o bien constituyeron, en realidad, un intento por establecer un punto de referencia de común acuerdo en cuyo marco la evaluación inicial no contaba con el respaldo pleno de las principales partes interesadas del país. En algunas evaluaciones también se trató de medir los progresos realizados desde la evaluación de 2004, utilizada para probar la versión preliminar del Marco del PEFA. Dado que dicha versión sufrió modificaciones considerables antes de convertirse en definitiva, no se puede realizar una comparación, salvo en el caso de ciertos indicadores.

10. Una segunda técnica cada vez más utilizada en los informes de evaluación del PEFA para realizar el seguimiento del desempeño en materia de GFP a lo largo del tiempo consiste en computar calificaciones “promedio” para las dimensiones fundamentales. Dado que estas calificaciones son más fáciles de computar a partir de números, los investigadores suelen convertir las calificaciones alfabéticas a numéricas y luego calcular un promedio aritmético, suponiendo ponderaciones iguales para cada indicador; todos los indicadores tienen la misma importancia.
11. Un ejemplo de ello es la evaluación del PEFA sobre Mozambique concluida y publicada en 2008. Las calificaciones en letras se convirtieron en calificaciones numéricas del 1 al 7 (D=1, D+=2), se agregaron para cada dimensión fundamental y se promediaron, suponiendo (sin explicación) ponderaciones iguales. La dimensión fundamental, que comprende la previsibilidad y control en la ejecución del presupuesto, se dividió en tres subdimensiones fundamentales: i) Recaudación y administración de ingresos: ID-13 a ID-15; ii) Gestión de efectivo: ID-16 a ID-17, y iii) Nómina, adquisiciones y controles internos (ID-18 a ID-21). Ello responde a la lógica de que estas áreas son conceptualmente distintas entre sí. Se agregó la categoría “Prácticas de los donantes” (D-1 a D-3).

Gráfico 1: Seguimiento de los progresos en Mozambique



Nota: En un principio, las calificaciones agregadas se expresaron en números (1; 1.5...4) y luego volvieron a expresarse en letras.

12. Sin perjuicio de las cuestiones relacionadas con el supuesto de ponderaciones iguales (que se analizan con mayor detalle más adelante), en el informe sobre Mozambique se demuestra que se puede hacer el seguimiento de los progresos a partir de una dimensión fundamental y en general. Uno de los principales objetivos era seguir de cerca los avances logrados desde la primera evaluación. Al establecer la dimensión fundamental y las calificaciones generales para cada evaluación, se pueden observar claramente las áreas en las que se registran avances (gráfico 1). Las diferencias en el largo de cada barra indican el alcance del progreso o retroceso. Cuando se observa únicamente el cambio, las cuestiones relacionadas con la validez del método de agregación se reducen, ya que la atención no se centra en los niveles; es decir, algunas de las calificaciones de las dimensiones fundamentales podrían ser diferentes si se calcularan a partir de una base diferente, por ejemplo agregando valores a partir de promedios ponderados y no de promedios aritméticos. De este modo, se supone que las cuestiones relacionadas con la agregación son las mismas en cada evaluación, lo que probablemente sea realista durante un período de apenas unos pocos años.

13. En la segunda evaluación del PEFA sobre Zambia, realizada a mediados de 2008, se utiliza la misma técnica (tampoco se explica la validez de suponer ponderaciones iguales); la única diferencia es que la dimensión fundamental relativa a la ejecución del presupuesto se divide en dos componentes (administración tributaria y ejecución del presupuesto), en lugar de tres.

3.2 Comparación entre países

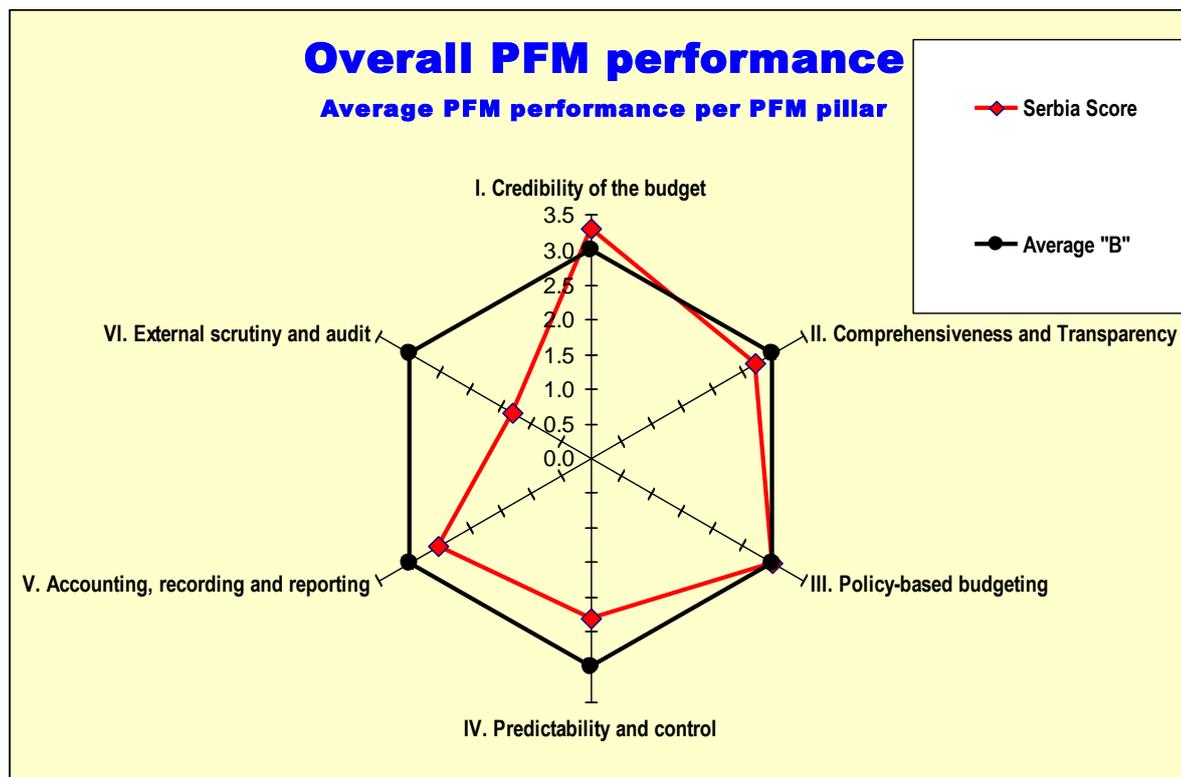
14. En el documento *Taking Stock: What do PEFA Assessments tell us about PFM systems across countries?* (Análisis: ¿Qué nos dicen las evaluaciones del PEFA acerca de los sistemas de GFP en los distintos países?) (Paulo De Renzio, Universidad de Oxford, abril de 2008), se utilizan datos sobre calificaciones del PEFA para buscar características de los países (como los ingresos per cápita y el tamaño de la población) que se encuentran correlacionadas a nivel mundial, y posiblemente para tratar de explicar las diferencias de desempeño en materia de GFP entre los países. En el documento se utiliza una conversión numérica de las calificaciones del PEFA y se asignan ponderaciones iguales a cada indicador y a todos los países de la muestra de 57 países.
15. Asimismo, se reconocen las cuestiones que puedan surgir en torno a la validez: “Existen varias razones por las que se podría criticar el uso de valores promedio; la principal es que las diversas dimensiones de la metodología del PEFA en realidad miden cosas muy diferentes, que no necesariamente se prestan a las conversiones cuantitativas, los cálculos o los análisis. Además, el uso de promedios se basa en el supuesto de que todos los indicadores tienen la misma importancia”. Sin embargo, el autor justifica dicho uso a partir de la gran muestra de países utilizada para el análisis y de la estricta metodología aplicada.
16. En otros casos, los informes de evaluación pueden incluir comparaciones del desempeño de un país con el desempeño “promedio” de otros países de la misma región, donde se ha aplicado la metodología del PEFA para evaluar la GFP, por ejemplo en forma de gráfico, como en el gráfico 2 que se muestra a continuación. Los supuestos formulados acerca de la conversión de calificaciones y las ponderaciones son similares al enfoque del informe que antes descrito.

3.3 Comparación simultánea entre países a lo largo del tiempo

17. Solo en unos pocos casos se llevaron a cabo evaluaciones reiteradas aplicando la metodología del PEFA con la finalidad expresa de hacer el seguimiento de los progresos en el desempeño a lo largo del tiempo. Por lo tanto, las calificaciones del PEFA para comparar los cambios en el desempeño regional o mundial a lo largo del tiempo solo se han utilizado de manera indirecta, o bien solo se ha establecido un punto de referencia que permite medir el cambio una vez que el número de evaluaciones reiteradas ha aumentado lo suficiente³. A continuación se presentan tres ejemplos de uso indirecto.

³ Ref.: Marco de resultados del BAfD.

Gráfico 2: Ejemplo de comparación del desempeño de un país con un valor de referencia regional



18. Las calificaciones de la evaluación de las políticas e instituciones nacionales del Banco Mundial (EPIN) se basan en evaluaciones del PEFA como fundamento de la calificación del indicador n.º 13 de la EPIN principalmente, aunque también, en cierta medida, del n.º 16 (sobre administración tributaria). No existe una conversión directa de las calificaciones del PEFA a las de la EPIN, pero se trata de garantizar la coherencia entre los dos conjuntos de calificaciones. Las calificaciones de la EPIN a su vez se utilizan para hacer el seguimiento de la Declaración de París a fin de medir las mejoras a nivel mundial en la calidad de los sistemas de GFP entre los países asociados del Comité de Asistencia para el Desarrollo.
19. En la Evaluación del apoyo al presupuesto general de 2006 se utilizaron los indicadores del PEFA como base para medir los cambios de desempeño en materia de GFP a lo largo del tiempo en los estudios de casos de países seleccionados que se utilizaron para extraer conclusiones a nivel mundial. Debido a la falta de evaluaciones reiteradas, los evaluadores se vieron obligados a realizar evaluaciones retroactivas de los indicadores como parte de los estudios sobre el terreno y aplicar una versión simplificada de la metodología de calificación del PEFA.
20. En el documento *Tracking Progress in the Quality of PFM Systems in HIPC*s (Seguimiento de los progresos en la calidad de los sistemas de GFP en los PPME) (P. De Renzio y W. Dorotinsky, 2007, Secretaría del PEFA, noviembre de 2007) se trata de hacer un uso más directo y detallado de las calificaciones del PEFA y de medir las tendencias mundiales del desempeño a lo largo del tiempo. Debido a la falta de evaluaciones reiteradas del PEFA, se utilizaron las calificaciones de los indicadores correspondientes a los planes de acción y evaluación de los países

pobres muy endeudados (PPME) como punto de referencia y calificaciones más recientes del PEFA como evaluaciones “reiteradas”. Luego fue necesario convertir las calificaciones del PEFA a calificaciones de los PPME. En el informe también se hace hincapié en cuestiones metodológicas y su posible impacto en los resultados.

4. Cuestiones derivadas de la agregación y la comparación

21. La agregación de los resultados correspondientes a las evaluaciones de los países entraña tanto la conversión de las calificaciones ordinales de los indicadores a valores numéricos como la asignación de ponderaciones a los distintos indicadores. Cuando se agrupan los resultados de los países, la agregación también requiere la asignación de ponderaciones a cada país. Cada una de estas cuestiones se analiza con mayor detalle a continuación.

4.1 Agregación de los distintos indicadores

Cuestión de las ponderaciones de los indicadores

22. La asignación de ponderaciones a los indicadores individuales resulta necesaria si los resultados de las evaluaciones se presentarán en términos más sencillos que 28(+3) resultados individuales o una redacción puramente descriptiva. En esos ejercicios se suelen asignar las mismas ponderaciones a los indicadores. Ahora bien, ¿se tiene una idea cabal de las consecuencias? Por ejemplo, en el cuadro 1 el cambio general pasaría de +1 a 1,5 puntos en la escala de calificaciones (una diferencia significativa) si el ID-11 y el ID-12 recibieran el doble de ponderaciones que los otros indicadores; esto último podría llegar a justificarse al incluir la dimensión fundamental de la GFP, “presupuestación basada en las políticas”, solo dos indicadores, y no de 3 a 8 indicadores, como las demás dimensiones fundamentales de la GFP.

23. Desde luego, el supuesto de ponderaciones iguales resulta muy conveniente, dado que facilita la agregación. Además, resulta muy provechoso para los usuarios de las calificaciones de evaluación del PEFA que deseen realizar agregaciones y obtener promedios a partir de subgrupos de indicadores distintos de las seis dimensiones fundamentales del PEFA (por ejemplo, el caso de Mozambique que se menciona más adelante y las calificaciones de la EPIN del Banco Mundial). Ahora bien, ¿es válido este supuesto?

24. El supuesto de ponderaciones iguales puede ser válido, aunque no siempre, de modo que los usuarios del Marco no deben suponer ponderaciones iguales automáticamente ni utilizar otros métodos numéricos de agregación en forma poco transparente y arbitraria (promedio ponderado o vínculo más débil entre los indicadores que se agregarán).

25. Los indicadores de desempeño 7, 8 y 9 del PEFA son particularmente controvertidos. La calificación D en la dimensión ii) del ID-7, que comprende la presentación de informes fiscales sobre la ayuda de los donantes recibida en efectivo, daría como resultado una calificación D o D+ para el indicador en general (método de calificación M1), pero puede revestir poca importancia si dicha ayuda representa solo un componente menor del gasto público. En el caso de los indicadores de desempeño 8 y 9, las calificaciones D pueden tener una importancia mucho mayor si el gasto de los Gobiernos subnacionales es un componente significativo del total del gasto público. Por lo tanto, en el caso de la dimensión fundamental relativa al alcance y la transparencia (ID-5 a ID-10), podría ser conveniente contar con una calificación basada en el promedio ponderado. Pero no

existe un criterio universal que permita determinar cuáles deberían ser esas ponderaciones.

26. Además, el supuesto de ponderaciones iguales para los indicadores de desempeño 17 a 21 no debería aplicarse en forma automática. En el caso del ID-17 (gestión de los saldos de caja y la deuda), las bajas calificaciones registradas en las dimensiones i) y iii) pueden tener muy poca repercusión en el funcionamiento del sistema de GFP en su conjunto si la deuda se destina a financiar solo una parte mínima del gasto público y las garantías de préstamos emitidas por el Gobierno son de poca magnitud. El ID-18 (gestión de la nómina) puede ser mucho más importante que los indicadores 19 a 21 (relativos a los controles del gasto no salarial) si la remuneración del personal es un componente importante del gasto público y el gasto no salarial es un componente menor. Una calificación baja en el caso del ID-21 (auditoría interna) puede revestir poca importancia si aún se está tratando de establecer un mecanismo de auditoría interna y se cuenta con sólidos sistemas de control interno *ex ante*. Por lo tanto, podría ser aconsejable aplicar un enfoque basado en el promedio ponderado a los indicadores de desempeño 13 a 21 (control y previsibilidad en la ejecución del presupuesto).
27. Sin embargo, la cuestión no es del todo clara. Por ejemplo, las deficiencias en los controles del gasto no salarial podrían afectar la calidad de los servicios públicos que presta el personal (un control deficiente de las adquisiciones de libros de texto podría menoscabar la eficacia de los docentes y, por ende, la eficiente prestación de servicios).
28. En el caso de los indicadores de desempeño 1 a 4 del PEFA (credibilidad de la dimensión fundamental relativa al presupuesto), podría argumentarse que convendría aplicar un enfoque basado en el vínculo más débil para obtener una calificación agregada aplicable a esta dimensión fundamental. No obstante, las circunstancias pueden variar y resulta difícil determinar a ciencia cierta qué método debería utilizarse. Por ejemplo, el hecho de que se asigne la calificación D al ID-2 y calificaciones A a los demás indicadores socava la credibilidad del presupuesto, dado que los ministerios, departamentos y organismos (MDO) están recibiendo recursos financieros muy distintos de lo presupuestado originalmente para ellos. El hecho de que se asigne la calificación D al ID-4 y calificaciones A a los demás indicadores podría indicar que ciertos pagos no se han incluido en el presupuesto. Ello implicaría la posibilidad de compensar futuras reducciones presupuestarias para los MDO en relación con lo que estos habían previsto, lo que afectaría la credibilidad del presupuesto. Por otra parte, la calificación D podría estar relacionada con grandes atrasos históricos que se van cancelando en forma gradual, aunque previsible.
29. Suponiendo que la agregación de calificaciones de acuerdo con las dimensiones fundamentales es válida, ¿es válido obtener una agregación general, de manera que una sola letra (o un solo número) indique el desempeño general en materia de GFP? La respuesta es complicada, debido a los vínculos bidireccionales que existen entre las dimensiones fundamentales y al hecho de que, por ejemplo, los indicadores de desempeño 1 a 4 se miden a lo largo de un período renovable de tres años y pueden reflejar el impacto del desempeño en los demás indicadores, la mayoría de los cuales se miden tomando como base el ciclo presupuestario de un año más reciente. Evaluar la importancia relativa de cada dimensión fundamental sería un ejercicio

difícil y arbitrario. En el caso de que se intente realizar ese tipo de agregación, el cálculo de promedios aritméticos puede ser el único método plausible, incluso si los vínculos no tienen la misma fuerza, pero es preciso ser consciente de las limitaciones de dicho método⁴.

30. Podría sostenerse que aplicar las mismas ponderaciones a cada dimensión fundamental no es siempre válido, dado que la cantidad de indicadores incluidos en cada dimensión fundamental varía ampliamente entre dos (ID-11 y ID-12⁵) y nueve (ID-13 a ID-21). Esto podría significar que la agregación directa a partir de indicadores individuales para llegar a un promedio general conduciría a un resultado distinto del que arrojaría una agregación realizada en dos pasos a través de las dimensiones fundamentales (las dimensiones 6 y 7 en el marco estándar, pero 9 en el caso de Mozambique, como se ilustró antes).

4.2 Conversión de una escala ordinal a una escala numérica

31. En la mayoría de los ejemplos de la sección 3, se utiliza la conversión de las calificaciones de la escala ordinal del PEFA (A, B, C, D) a valores numéricos, salvo cuando solo se utilizan distribuciones de calificaciones. Comúnmente, las cuatro calificaciones ordinales pasan a ser (4,3,2,1) o (3,2,1,0)⁶ y se asigna un valor de medio punto al signo “+” (por ejemplo, B=3, B+=3.5).
32. Una cuestión fundamental consiste en determinar si es válido suponer que existen las mismas diferencias de calidad entre cada clasificación y que los cambios en las calificaciones significan lo mismo para todas las calificaciones iniciales. Dado que la mayoría de las calificaciones se basa en diferentes pasos para mejorar los componentes de procesos, uso de sistemas y productos descritos en forma cualitativa, no existe una base científica para hacer la conversión a valores numéricos⁷. Dejando de lado los casos “sin calificación” (mencionados más adelante), en la investigación dirigida por el Grupo de Investigaciones sobre el Desarrollo y la Unidad de Gestión del Sector Público del Banco Mundial se señala, sin embargo, que las variaciones moderadas en la escala de conversión numérica probablemente no den como resultado conclusiones muy distintas.

Cuestión de los métodos de calificación M1 y M2

33. A primera vista, también parecería ser inválido derivar calificaciones numéricas de las dimensiones fundamentales cuando se asignan calificaciones a los indicadores

⁴ Un ejercicio útil podría ser calcular los coeficientes de correlación entre las calificaciones promedio de las dimensiones fundamentales 2 a 6 tomadas en conjunto (es decir, el promedio de la suma de las calificaciones asignadas a estas dimensiones) y la primera dimensión. Tal como se muestra en el documento del Marco, la primera dimensión fundamental (credibilidad del presupuesto) se define como el resultado del impacto combinado de las otras dimensiones fundamentales. Sin embargo, un coeficiente de correlación bajo no significa necesariamente que exista falta de correlación entre las demás dimensiones fundamentales; el hecho de que la primera reciba calificaciones altas y se asignen calificaciones bajas a las demás puede indicar simplemente posibles amenazas a la credibilidad del presupuesto. En una buena evaluación resumida se debería señalar este aspecto.

⁵ Téngase en cuenta que los aspectos relacionados se incluyen en otras dimensiones, por ejemplo los ID-5 a ID-7, el ID-10 y el ID-27.

⁶ El hecho de que se opte por la primera o la segunda conversión no influye decisivamente en los resultados de la agregación, a menos que también se asigne un valor a la categoría “sin calificación”, como se muestra más adelante.

⁷ Los indicadores meramente cuantitativos ID-1 a ID-3 y D-3 podrían considerarse excepciones.

según los métodos M1 (vínculo más débil) y M2 (promedio aritmético). La calificación D+ no siempre significa exactamente lo mismo cuando se asigna al indicador M1 que cuando se trata del indicador M2, ya que se utilizan distintos procesos de cálculo. La cuestión consiste en determinar si las diferencias son lo bastante significativas como para merecer un grado de atención considerable. Hasta el momento, ningún usuario ha considerado que la cuestión fuera lo suficientemente importante como para justificar una conversión diferente a valores numéricos.

La cuestión de los casos “sin calificación”

34. La validez de la conversión de calificaciones alfabéticas a numéricas y la posterior agregación también se ve afectada por la incidencia de los casos “sin calificación”, debido a la falta de información adecuada o a problemas para acceder a ella, o a indicadores que no reciben ninguna calificación porque se consideran inaplicables (o se los excluye deliberadamente de la evaluación). En este caso, la agregación y el cálculo de promedios puede dar lugar a resultados engañosos. Si un país no dispone de datos sobre indicadores que probablemente reciban calificaciones altas (A o B), la calificación agregada mostraría un sesgo por defecto, y sucedería lo contrario cuando faltan datos sobre indicadores que probablemente reciban calificaciones bajas.
35. Cabe preguntarse entonces si los casos “sin calificación” deberían recibir un valor a los fines de la agregación y si deberían tenerse en cuenta en la calificación promedio o la distribución de las calificaciones. Hasta el momento, han quedado excluidos de la agregación en la mayoría de los ejemplos. Puede sostenerse, sin embargo, que la inexistencia de calificación producto de la falta de datos adecuados representa una situación más desfavorable que la calificación D, pues implica que ni siquiera se dispone de la información básica para evaluar la situación. En algunos ejemplos reales de agregación se tiene en cuenta la categoría “sin calificación”. El BAfD, por caso, la incluye en su marco institucional de resultados asignándole el valor 0 (cero) dentro de una escala en la que $D=1$ y $A=7$. Como resultado, se obtiene una calificación agregada que difiere de un sistema de calificación en que la inexistencia de calificación vale 1 o directamente no se computa.
36. En la base de datos de las calificaciones correspondientes a los informes de evaluación, que se encuentra en la Secretaría del PEFA, se hace una distinción entre las distintas razones a favor de la categoría “sin calificación” subdividiéndola en NC (no calificada por falta de información), NA (no aplicable en el contexto específico del país) y NU (no utilizado para la evaluación; se trata de una decisión *ex ante* para limitar el alcance de la evaluación). Estas tres categorías deberían incorporarse de distinto modo a la hora de agregar resultados de desempeño y la categoría NC sería la única que podría llegar a compararse o agregarse con otras calificaciones.

Distribución de frecuencias

37. Otro método de agregación, que permite evitar la conversión numérica, consiste en compilar en forma de gráficos las distribuciones de frecuencias de las calificaciones: número de calificaciones D como porcentaje del total de calificaciones, número de calificaciones C,... número de calificaciones A. Las categorías A y D suelen comprender un porcentaje menor de calificaciones que las categorías B y C, de modo que en dichos gráficos se observaría una “cola” en ambos extremos. Los

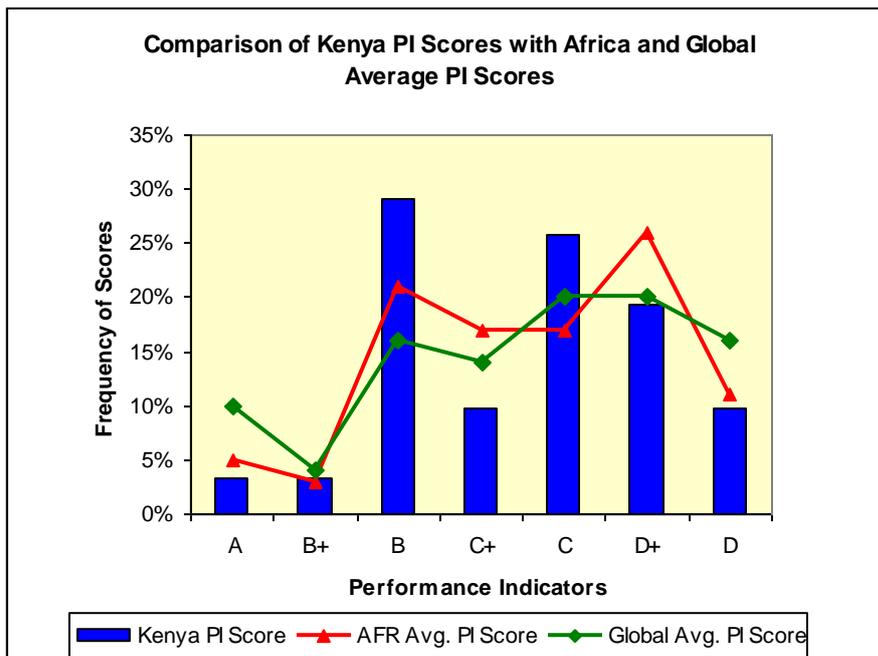
cambios en el funcionamiento del sistema de GFP de un país a lo largo del tiempo pueden evaluarse comparando las distribuciones de frecuencias en un gráfico: es de esperar que la proporción de calificaciones A y B se incremente con el tiempo.

38. Si bien el uso del método de distribución de frecuencias permite evitar las conversiones numéricas, algunas de las cuestiones relativas a la validez siguen siendo las mismas. El supuesto implícito es que todas las calificaciones A (por ejemplo) tienen el mismo valor, es decir que todos los indicadores tienen las mismas ponderaciones, y que a un determinado nivel de calificaciones el desempeño es igual —o al menos comparable— en todos los indicadores. Una desventaja específica relacionada con el método de conversión numérica es que resultaría difícil mostrar en un gráfico la distribución de frecuencias de cada dimensión fundamental en distintos momentos; en el gráfico 1 se muestra que ello es posible, aunque solo si se aplica un método de conversión numérica.
39. Si bien las presentaciones de distribuciones de frecuencias en general muestran las calificaciones como si estas tuvieran intervalos del mismo valor, la presentación de las frecuencias de las calificaciones también puede utilizarse para ilustrar el vínculo que existe entre una calificación de base (por ejemplo, una B) y su contrapartida con “+” (B+), mostrándolas en forma consecutiva, con un intervalo al siguiente grupo (por ejemplo, C y C+).

4.3 Agrupación de países

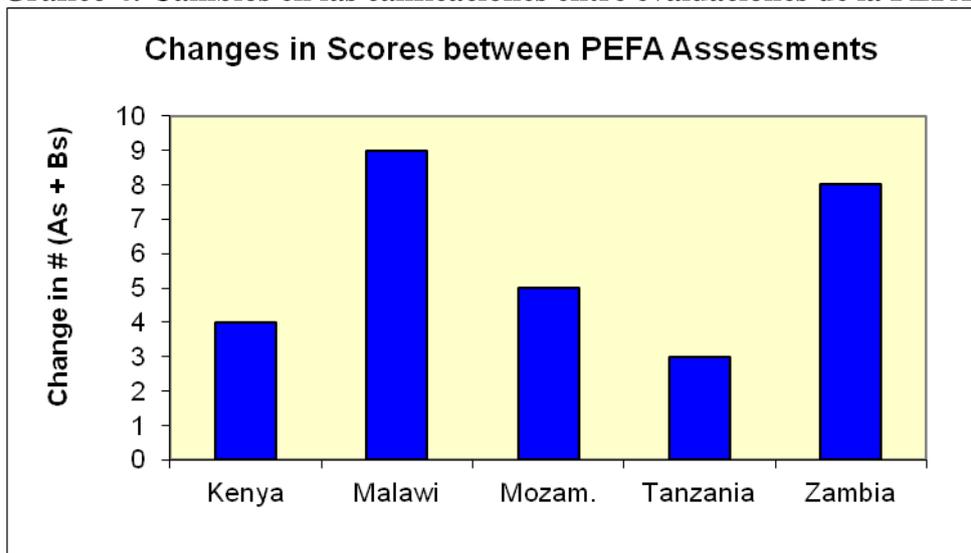
40. Los métodos de agregación numérica descritos anteriormente también pueden utilizarse para facilitar la comparación entre países, tanto en un momento determinado como a lo largo de distintos períodos. Utilizando el método de conversión numérica, se podría modificar el gráfico 1 para mostrar las calificaciones agregadas generales correspondientes a diferentes países o grupos de países, cada uno de ellos representados por una barra. Los grupos podrían representar regiones o grupos de países con características similares. También podría modificarse el gráfico para mostrar las calificaciones agregadas generales de un país frente a un promedio regional o mundial. Otra opción sería mostrar las calificaciones promedio de cada dimensión fundamental para todos los países o una muestra de ellos. No obstante, si bien es difícil mostrar en un gráfico comparaciones entre países por dimensión fundamental, podrían incluirse varios gráficos (uno de las calificaciones agregadas generales y otro de cada dimensión fundamental) en una página (véase también el párrafo 45).
41. El método de agregación basado en la distribución de frecuencias puede usarse prácticamente de la misma manera. Por ejemplo, en el gráfico 3 se compara la distribución de frecuencias de las calificaciones de un país africano (Kenya) con la de toda África y la de todos los países del mundo.

Gráfico 3: Distribución de frecuencias de las calificaciones: Comparación de Kenya con el promedio mundial y regional



42. Se pueden mostrar más países en un gráfico mediante la combinación de calificaciones, por ejemplo agrupando las calificaciones A y B, por un lado, y las C, D y NC, por el otro. Esto permite apreciar más claramente la diferencia de desempeño en materia de GFP entre los países.
43. Mediante este método también se pueden seguir de cerca los progresos a lo largo del tiempo, mostrando los cambios en el número de calificaciones A y B combinadas a la luz de una evaluación reiterada. Por definición, el aumento del número de A y B combinadas es idéntico a la reducción del número de C, D y NC combinadas, por lo que no es necesario mostrar también estas últimas. En el gráfico 4 se proporciona un ejemplo.

Gráfico 4: Cambios en las calificaciones entre evaluaciones de la PEFA



i) fechas de las evaluaciones: Kenya: 2006; Malawi: 2006 y 2008; Mozambique: 2006 y 2007; Tanzania: 2006; Zambia: 2005 y 2008.

44. En el gráfico 4 se han utilizado calificaciones reales de evaluaciones del PEFA para las evaluaciones reiteradas realizadas en Malawi y Mozambique⁸. En los casos de Kenya, Tanzania y Zambia, se han utilizado calificaciones meramente ilustrativas de las evaluaciones más recientes. En el gráfico 4 se observa que Malawi es el país con más aumentos en el número de A y B, seguido de Zambia y Mozambique. El gráfico también podría servir para comparar el desempeño de las distintas regiones en materia de GFP, a cuyos efectos cada barra representaría una región.
45. La distribución de frecuencias de las calificaciones también puede utilizarse para comparar las calificaciones asignadas a un grupo de países en cada dimensión fundamental o incluso a indicadores específicos. Se pueden compilar ocho gráficos en una o dos páginas para mostrar la evolución general de la reforma y los avances de esta en cada una de las seis dimensiones fundamentales. De este modo, se puede ver a grandes rasgos en qué medida estos avances contribuyen al progreso general.

Cuestiones derivadas de las comparaciones entre países

46. La comparación entre países genera cuestiones en numerosas áreas socioeconómicas y los indicadores del desempeño en materia de GFP no son la excepción. Existe el riesgo de que los elementos de la comparación no sean estrictamente comparables. Por ejemplo, las diferencias en la metodología estadística y los problemas en la recopilación de datos significan que es necesario actuar con precaución a la hora de comparar las estimaciones relativas al producto interno bruto (PIB) entre países. La eficacia de las comparaciones entre países también puede verse afectada cuando existen diferencias en las definiciones. Por ejemplo, el gasto público en servicios de salud de los países suele compararse en términos de porcentajes del gasto y del PIB, sin tener debidamente en cuenta las diferencias que existen entre las definiciones del

⁸ Se han tomado como base los informes finales de Malawi, Mozambique y Zambia. Los informes de Kenya y Tanzania no estaban concluidos al momento del cálculo. Una vez que estos se hayan finalizado y que se haya agregado a Uganda (la nueva evaluación realizada a fines de 2008 está a punto de concluir), el gráfico 4 podría mostrar un panorama interesante.

sector de salud (por ejemplo, el abastecimiento de agua y saneamiento se incluye en algunos países y se excluye en otros), las diferencias en cuanto al grado de participación del sector privado en la prestación de servicios de salud y las diferencias en las condiciones sanitarias.

47. Aunque se puedan estimar válidamente las calificaciones agregadas de las seis dimensiones fundamentales de la GFP, la comparación de los indicadores del PEFA entre distintos países generará diversos problemas de disparidad, entre ellos los siguientes:

- El alcance (el Gobierno central, los Gobiernos subnacionales o ambos) y el año de la evaluación pueden no ser los mismos para los países que se comparan.
- Las definiciones de los términos de la GFP (por ejemplo, atraso en los pagos, organismos públicos o sistemas de transferencia fiscal) pueden variar de un país al otro.
- Las razones para asignar a distintos países las mismas calificaciones a cada indicador pueden variar considerablemente, dado que la mayoría de los indicadores consta de al menos dos y a veces cuatro dimensiones. Por ejemplo, dos países pueden obtener la misma calificación en el ID-8, aunque por razones totalmente distintas: mientras que uno recibe una A en la dimensión 1 y una D en la dimensión 3, otro recibe una D en la dimensión 1 y una A en la dimensión 3. Por lo tanto, las comparaciones entre países probablemente sean superficiales y deban hacerse a nivel de cada dimensión, lo que resulta un ejercicio más complicado.
- La calidad de las evaluaciones puede variar de un país a otro. En el proceso de garantía de la calidad a cargo de la Secretaría (y otros expertos evaluadores), se identifican deficiencias en los informes de evaluación del PEFA. En el caso de la Secretaría, la deficiencia consiste principalmente en la existencia de calificaciones de indicadores no respaldadas por pruebas suficientes (caso más común) y de calificaciones incorrectas evaluadas sobre la base de la información suministrada. Los informes revisados a partir de los comentarios recibidos de expertos evaluadores no necesariamente se envían a la Secretaría para que esta realice un nuevo examen⁹. De este modo, los informes finalizados pueden contener errores en las calificaciones.
- Las razones para asignar ponderaciones a los indicadores para calcular una calificación agregada de la dimensión fundamental o una calificación general (promedios aritméticos —a partir del supuesto de igualdad de ponderaciones—, promedios ponderados o enfoques basados en el vínculo más débil) muy probablemente variarán de un país al otro, como se explicó antes.

⁹ Secretaría del PEFA: *Report on Early Experience of the Application of the Framework* (Informe sobre las primeras experiencias de aplicación del Marco), noviembre de 2006, y *PFM Performance Measurement Framework Monitoring Report* (Informe de seguimiento del Marco de Medición del Desempeño de la GFP), marzo de 2008, www.pefa.org.

- Existen diferencias entre los países en cuanto a la proporción del gasto público financiada directamente por el Gobierno y la proporción financiada directamente por los donantes. Por ejemplo, la comparación de las calificaciones del PEFA entre un país donde el 50% del gasto público es financiado directamente por los donantes y otro país donde dicho valor es del 0% posiblemente sea menos significativa que la comparación de las calificaciones del PEFA entre países con porcentajes similares.
- La incidencia de los indicadores “no calificado”, “no utilizado” y “no aplicable” puede variar de un país al otro, lo que reduce la posibilidad de comparar los promedios de las calificaciones, dado que tal vez no siempre se hayan calificado los mismos indicadores.
- Otra cuestión consiste en determinar si las calificaciones de los países deben ponderarse según el tamaño de la población, por ejemplo. Lo más probable es que esta cuestión surja dentro de un grupo regional de países. Un país muy pequeño dentro de la región puede obtener la calificación D en todas las categorías, lo que reduce, en el caso de que se utilicen ponderaciones iguales, la calificación promedio de la región (obtenida a través del método de conversión numérica) si los demás países de mucho mayor tamaño reciben principalmente calificaciones A y B. El método elegido podría afectar la comparabilidad entre distintos grupos de países.

4.4 Formas de reducir los problemas de comparabilidad

Tamaño de la muestra

48. Cuanto mayor es la muestra, menor es el riesgo de comparaciones inválidas. Los mayores tamaños de muestras se observarían: i) cuando se comparan entre sí las calificaciones obtenidas en las dimensiones fundamentales del PEFA por todos los países donde se han realizado evaluaciones (es decir, si el número asciende a 50, la calificación de cada dimensión se calcula como el promedio de las calificaciones de los 50 países), y ii) cuando se compara un país con el promedio de todos los demás países (véase el gráfico 3 anterior), suponiendo que se trata de un país razonablemente representativo.
49. La muestra más pequeña para comparar las calificaciones obtenidas en las dimensiones fundamentales del PEFA se observaría cuando la comparación se realiza entre los países individualmente; en tal caso, podría llegar a cuestionarse la validez de dichas comparaciones. Una manera de aumentar el tamaño de la muestra sería dividir los países en grupos. Sin embargo, los grupos solo incluirán un número limitado de países y pueden surgir cuestiones en torno a la validez de la comparación¹⁰.

Comparaciones entre países con características similares

¹⁰ Documento de De Renzio, página 10; las comparaciones se presentan en forma de gráfico tubular o de barras.

50. Entre los ejemplos se incluyen los países de una determinada región geográfica de la ex Unión Soviética (por ejemplo, el Cáucaso), los países angloparlantes de una determinada región de África (por ejemplo, África oriental) y los países angloparlantes del Caribe, que presentan características en común, tienen regímenes jurídicos e institucionales similares y han emprendido, en líneas generales, el mismo camino de reforma de la GFP (incluso en el marco del mismo programa de asistencia técnica respaldada por donantes).

Prestar especial atención a los cambios en el desempeño en materia de GFP

51. La cuestión del tamaño de la muestra también se torna menos importante si se comparan los cambios en el desempeño en materia de GFP a lo largo del tiempo, antes que el nivel de dicho desempeño en un momento determinado. Si se supone que los problemas específicos de cada país se mantienen relativamente invariables en el tiempo (probablemente un supuesto razonable), la comparación de los cambios en el desempeño en materia de GFP elimina la influencia de dichos problemas. Sin embargo, en los informes que incluyan comparaciones entre países habrá que demostrar que los factores específicos del país no se han modificado con el tiempo y que, por lo tanto, los cambios registrados en las calificaciones a lo largo del tiempo revisten la misma importancia (en términos del cambio en la calidad del sistema de GFP) para todos los países incluidos en la comparación.

5. Conclusiones

52. Uno de los principales objetivos del Marco del PEFA es comparar las calificaciones de las evaluaciones llevadas a cabo en un país determinado a lo largo del tiempo. Cuando la comparación se realiza por indicador y va acompañada de un resumen pormenorizado, no genera problemas metodológicos, aunque puede complicarse la tarea de brindar a quienes no son especialistas información sencilla acerca de las tendencias en materia de desempeño. Toda agregación de las calificaciones de los indicadores destinada a proporcionar medidas de calificaciones sencillas plantea cuestiones relacionadas con la conversión de la escala numérica y las ponderaciones de los indicadores.
53. La comparación del desempeño de los países en materia de GFP a partir de las calificaciones obtenidas en las evaluaciones del PEFA es válida en principio, pero puede presentar riesgos. Los analistas deberían destacar expresamente las cuestiones que puedan menoscabar la validez de dichas comparaciones y justificar cualquier supuesto formulado al respecto (por ejemplo, el uso de ponderaciones iguales para los indicadores).
54. En todos los informes en que se aplican técnicas de agregación —encontrados hasta el momento—, se ha supuesto que todos los indicadores tienen igual ponderación y que los intervalos numéricos entre las calificaciones alfabéticas de la escala ordinal son de igual magnitud.
55. Otra manera de realizar una comparación entre países es comparar la distribución de las calificaciones (cantidad de A, B, C y D) de estos en un determinado momento y a lo largo del tiempo. Este método brinda un análisis potencialmente más detallado y evita la conversión de calificaciones alfabéticas a numéricas.
56. Sin embargo, ambos métodos plantean cuestiones que deberían tenerse en cuenta al hacer comparaciones entre países. Además, surgen otras cuestiones. En ninguno de los informes encontrados hasta el momento se analizan claramente los supuestos formulados, sus fundamentos ni el grado de sensibilidad de las conclusiones extraídas a propósito de esos supuestos. En todos ellos se da por sentado que los intervalos entre las calificaciones A, B, C y D son de igual magnitud, que todos los indicadores tienen la misma ponderación y que a todos los países les corresponde la misma ponderación dentro de los grupos. La ventaja de estos supuestos es que son sencillos, transparentes y susceptibles de reproducción.
57. No existe un método científicamente correcto para determinar cómo debería realizarse la agregación en cada uno de esos tres niveles de agregación. En consecuencia, el PEFA no respalda la agregación de resultados en general, así como ningún método de agregación en particular.
58. El PEFA recomienda a los usuarios que —como parte de la difusión de los resultados obtenidos en las comparaciones— expliquen claramente el método de agregación aplicado a cada caso. También sería aconsejable que los usuarios se sometieran a análisis de sensibilidad para destacar el grado de solidez de sus conclusiones a la luz de supuestos de agregación alternativos.

59. La validez de las comparaciones entre países aumenta según los siguientes aspectos:

- *El tamaño de la muestra cuando se agrupan países.* Cuanto mayor es la muestra, menos importantes se tornan las cuestiones de disparidad, dado que estas últimas son limitadas en número. La muestra más grande se observa cuando se comparan las calificaciones de las dimensiones fundamentales/calificaciones generales de todos los países, o cuando se compara un país con el promedio de todos los demás países (siempre que el país en cuestión no presente aspectos específicos que dificulten la comparación). La muestra más pequeña se observa cuando se compara un país con otro.
- *El número de semejanzas entre los países que se comparan:* La comparación de las calificaciones del PEFA obtenidas por países con características similares (como los niveles de ingreso, el tamaño o el legado administrativo) puede ser más legítima que la comparación de calificaciones de países que son muy distintos. Así, los países que integran grupos regionales homogéneos pueden compararse con mayor validez que los grupos regionales que se comparan con otros grupos regionales.
- *La medición de los cambios a lo largo del tiempo, antes que el nivel de desempeño absoluto:* La comparación de los cambios en las calificaciones de un país específico (o de un grupo de países) puede reducir el riesgo de comparaciones inválidas si los factores específicos del país en cuestión no cambian significativamente con el tiempo.
- *El grado de semejanza entre los países en cuanto a la incidencia y distribución de los casos “sin calificación”.* La validez de las comparaciones entre países se ve comprometida por la gran incidencia de los casos “sin calificación” y las diferencias entre los países relativas a la distribución de estos casos en el conjunto de indicadores, así como el grado de probabilidad de la prueba, si estuviese disponible, que avala las calificaciones altas y las bajas.

60. La conversión numérica de las calificaciones recibidas en las evaluaciones del PEFA es sin duda la manera más sencilla de realizar comparaciones entre países, pero, como se explica en este documento, da lugar a numerosas cuestiones de validez. Todavía queda por ver si estas cuestiones revisten la importancia suficiente para justificar la aplicación de los métodos más complicados para realizar comparaciones entre países, tal como se describe en forma teórica en la sección 3.

61. Por lo tanto, dado que la agregación parece estar abriéndose paso entre varias partes interesadas para diversas finalidades, puede resultar útil hacer un seguimiento del análisis teórico contenido en este documento investigando los efectos prácticos de los distintos enfoques en materia de agregación. Esto podría lograrse realizando un estudio para determinar en qué medida las consideraciones formuladas en este documento darían lugar a diferentes conclusiones en la comparación entre países o bien comparando los avances registrados en un país (o un conjunto determinado de países) a lo largo del tiempo, cuando se aplica a datos reales extraídos de la base de datos de las calificaciones correspondientes a las evaluaciones del PEFA.

APÉNDICE 1 Bibliografía

Fondo Africano de Desarrollo: *Results Reporting for ADF-10 and Results Measurement Framework for ADF-11*, documento de antecedentes: febrero de 2008.

Dorotinsky, William y De Renzio, Paulo: *Tracking Progress in the Quality of PFM Systems in HIPC Countries*; Secretaría del PEFA, noviembre de 2007.

De Renzio, Paulo: *Taking Stock: What do PEFA Assessments tell us about PFM systems across countries?*, Universidad de Oxford, abril de 2008.

Grupo de Evaluación Independiente, Banco Mundial: *Public Sector Reforms: What Works and Why? An IEG Evaluation of World Bank Support*, junio de 2008.

Kaiser, Kai y Steinhilper, David: *Note on PEFA and Fiscal Flows to Sub-National Governments and Front-Line Providers* (informe preliminar); Banco Mundial, 2008.

OCDE: *Methodology for assessment of national procurement systems*, versión 4, julio de 2006.

Programa de PEFA: Marco de Medición del Desempeño en materia de GFP, junio de 2005

Informes de evaluación del PEFA: Kenya, Malawi, Mozambique, Tanzania, Zambia; www.pefa.org.

Secretaría del PEFA: *Common Approach to PEFA Value Added and links to other PFM indicators*, agosto de 2007.

Secretaría del PEFA: *Report on Early Experience of the Application of the Framework*, noviembre de 2006, www.pefa.org.

Secretaría del PEFA: *PFM Performance Measurement Framework Monitoring Report*, marzo de 2008.

Banco Mundial: Evaluación de las políticas e instituciones nacionales (EPIN): para consultar acerca de la metodología, visite www.worldbank.org.