



Cadre de mesure de la performance de la GFP

Rapport d'étape 2010

*Analyse des évaluations répétées et de l'évolution de la performance des
systèmes de GFP mesurée au moyen des indicateurs PEFA*

Secrétariat PEFA

Rapport final

19 mai 2011

Table des matières

Sigles et abréviations	4
Résumé analytique	5
Chapitre 1 – Introduction et méthodologie	11
Chapitre 2 - Fréquence des ER et ses déterminants	16
2.1 Présentation générale	16
2.2 Observations.....	16
Chapitre 3 – Efficacité de la mesure de l'évolution	21
de la performance	21
3.1 Aperçu.....	21
3.2 Constatations.....	21
Chapitre 4 – Évolution de la performance de la GFP	33
4.1 Aperçu.....	33
4.2 Constatations.....	33
Chapitre 5 - Recommandations	42
Annexe A : Évaluations PEFA comparatives	44
Annexe B : évaluations PEFA répétées non considérées comme « comparatives »	45
Annexe C : Évolution possible des notes attribuées aux composantes	47
entre l'évaluation préalable et l'évaluation répétée	47
Annexe D : Pays ayant achevé une évaluation de référence avant le 30 juin 2006 (c'est-à-dire il y a plus de 4 ½ ans) mais qui n'ont pas encore achevé d'évaluation répétée	48
Annexe E : Pourcentage des composantes d'indicateur pouvant faire l'objet de comparaisons fiables entre EC, par évaluation, lorsque les mentions « non notés » sont exclues	49
Annexe F : Pourcentage des notes pouvant faire l'objet de comparaisons fiables entre EC, pour chaque composante d'indicateur, lorsque les mentions « non notés » sont exclues	50
Annexe G : Ventilation en pourcentage des indicateurs composantes des EC comparables, non notés et non comparables	51
Annexe H : Ventilation en pourcentage des changements des EC par indicateur et composante	52
Annexe I : Méthodologie de codage de la GFP	54
(établie par M. Andrews)	54
Annexe J : Bibliographie	56

Les principaux membres de l'équipe du rapport étaient Helena Grandão Ramos (coordinatrice de la mission), Clay Wescott et Brandon Lundberg. Les conseils techniques ont été fournis par Frans Ronsholt, Phil Sinnett, Charles Seibert et Tony

Bennett. L'examen par les pairs du rapport a été assuré par Paulo de Renzio (indépendant), Rachel Perrin (DFID), Monica Rubiolo (SECO) et Steve Knack (Banque mondiale). L'étude a été supervisée par Frans Ronsholt.

Sigles et abréviations

AusAid	Australian Agency for International Development (Agence australienne pour le Développement international)
BaFD	Banque africaine de développement
BAfD	Banque asiatique de développement
BID	Banque interaméricaine de développement
Cadre	Cadre de mesure de la performance de la gestion des finances publiques
CE	Commission européenne
CFAA	Évaluations de la responsabilité financière du pays
CIFA	Évaluation fiduciaire intégrée du pays
D-1, 2 ou 3	Indicateurs des pratiques des bailleurs de fonds
DFID	Department for International Development (Ministère britannique du développement international)
EC	Évaluation comparative
EP	Évaluation précédente
ERPFM	Examen externe de la gestion des finances publiques
FMI	Fonds monétaire international
GFP	Gestion des finances publiques
NdS	Note de synthèse
NdS/LdM	Note de synthèse /Lettres de mission
Norad	Agence norvégienne de développement international
OCDE-BIS	Organisation de coopération et de développement économique – ensemble d’indicateurs de référence de la passation des marchés publics
PEFA	Dépenses publiques et responsabilité financière
PEMFAR	Programme d’examen de la gestion des dépenses publiques et de la responsabilité financière
PER	Examen des dépenses publiques
PI	Indicateur de performance
RP-GFP	Rapport sur la performance de la gestion des finances publiques
SECO	Secrétariat d’État à l’économie (Suisse)

Résumé analytique

1. Le Cadre PEFA a été principalement conçu pour créer un outil de suivi permettant de mesurer l'évolution de la performance des systèmes de GFP dans le temps au niveau national. Ce quatrième rapport d'étape sur le déploiement du Cadre PEFA vise à déterminer si cet objectif a été atteint, et examine l'évolution des systèmes de GFP dans le monde par rapport à une première série d'évaluations répétées.
2. Le rapport couvre les rapports d'évaluation qui avaient été finalisés (rapports définitifs) ou achevés pour l'essentiel (version préliminaire intégrale du rapport) au 6 octobre 2010 et qui constituaient, pour les pays considérés, la deuxième ou la troisième série d'évaluations PEFA (dites évaluations répétées). Quarante-cinq évaluations répétées, qui concernent 38 pays, ont été effectuées.
3. Trente-trois de ces évaluations répétées (couvrant 29 pays) sont qualifiées d'« évaluations comparatives » (EC) parce qu'elles avaient pour objet de mesurer les changements dans le temps. Douze autres évaluations répétées ne rentrent pas dans cette catégorie car elles n'avaient pas pour but de mesurer l'évolution survenue depuis une évaluation antérieure.

Fréquence et raisons d'être des évaluations répétées

4. Cinq ans et quatre mois après le lancement du Cadre PEFA, 206 évaluations PEFA ont été effectuées dans 119 pays. Dans un tiers environ de ces pays, une évaluation répétée a également été menée. Dans un quart des 119 pays, l'évaluation répétée était une évaluation comparative.
5. Les évaluations comparatives ont été effectuées entre un et quatre ans après l'évaluation préalable (EP) (soit en moyenne 33 mois après cette évaluation). Sept des pays considérés ont mené une évaluation comparative moins de 30 mois après l'évaluation préalable, c'est-à-dire avant la fin du délai minimal recommandé qui est de 3 ans. Le délai moyen pour ces sept EC était de 22 mois.
6. Les motifs indiqués par les 26 pays qui ont procédé à une évaluation comparative 30 mois ou plus après l'évaluation préalable sont, principalement, la volonté de mesurer les progrès et de faciliter la conception ou le suivi d'un plan d'action ou d'un programme de réforme de la GFP et, dans une moindre mesure, le désir de faciliter le dialogue avec les bailleurs et l'établissement de liens avec des opérations d'appui au budget général en cours ou prévues. Les évaluations menées à plus brève échéance avaient essentiellement les mêmes motifs, mais ont été fondamentalement effectuées parce que l'assistance future des bailleurs était subordonnée aux résultats de l'évaluation comparative, ce qui explique l'urgence d'y procéder.
7. Les douze évaluations répétées qui n'ont pas été menées à des fins de comparaison ont été effectuées en moyenne 30 mois après l'évaluation précédente. Elles visaient principalement à établir des références, ce qui suggère que les principales parties prenantes n'avaient pas confiance de l'évaluation préalable.

8. Il n'a pas été possible de détecter l'existence d'une relation quelconque entre les raisons pour lesquelles des évaluations comparatives ont été effectuées, et leur fréquence, d'une part, et les caractéristiques des pays ou les bailleurs les parrainant, d'autre part.

9. Cinq pays n'avaient pas encore procédé à une évaluation comparative quatre ans et demi ou plus après leur dernière évaluation, mais deux d'entre eux prévoyaient ou avaient entrepris une évaluation répétée. Ces longs délais tiennent à des changements de gouvernement, aux retards pris par le lancement de programmes de réformes publiques, et à la difficulté que présente la réalisation d'une évaluation de référence lorsque l'adhésion des autorités est minimale ou inexistante.

10. Il est recommandé de répéter les évaluations PEFA entre trois et cinq ans après l'évaluation de référence. Globalement, 80% environ des évaluations comparatives ont été mises en œuvre après plus de 2 ans et demi. Cinq pays seulement n'avaient pas encore répété leur évaluation 4 ans et demi après l'évaluation initiale. Il semble donc que les évaluations répétées soient poursuivies de manière systématique par les pays, et le plus souvent dans les délais recommandés.

Efficacité de la mesure de l'évolution de la performance

11. Afin de pouvoir utiliser les notes des indicateurs PEFA pour mesurer l'évolution de la performance dans le temps, il importe d'être raisonnablement certain que chaque paire constituée par les notes attribuées à un indicateur (et à chacune de ses composantes) dans l'évaluation préliminaire et dans l'évaluation comparative représente l'évolution « réelle » de la performance. Le Secrétariat PEFA a examiné chaque composante des indicateurs de chacune des 33 évaluations comparatives pour s'assurer que les éléments concrets disponibles étaient suffisants pour attribuer une note dans le cadre tant de l'EP que de l'EC et aussi pour identifier les éventuels problèmes de notation (tels que la présentation de nouveaux éléments pour la notation de l'évaluation préalable, l'utilisation de définitions ou de méthodes d'échantillonnage différentes ou encore des interprétations divergentes à partir de données similaires).

12. La présence d'indicateurs « non notés » dans les évaluations préalables et/ou dans les évaluations comparatives compromet également la détermination de l'évolution de la performance dans le temps. Mais comme les cas de composantes d'indicateurs « non notés » sont aisément détectables lorsqu'on compare les paires de notes, les utilisateurs des données savent dans quel cas le problème se pose.

13. Lorsque les indicateurs ou leurs composantes « non notés » sont exclus des séries de données, un degré de comparabilité de 80 % ou plus entre les notes des EP et des EC pour toutes les composantes des indicateurs est jugé satisfaisant. Ce degré élevé de comparabilité a été atteint dans le cas de 25 des 33 évaluations comparatives, c'est-à-dire 76 % des rapports. Le degré de comparabilité a été jugé insuffisant (en ce sens que moins de la moitié des notes attribuées aux composantes des indicateurs peuvent faire l'objet de comparaisons fiables) pour 3 des 33 évaluations comparatives (9 %).

14. L'analyse des facteurs inhérents au processus susceptibles de contribuer à la robustesse de la mesure de l'évolution de la performance dans le temps a abouti à la conclusion que tous les facteurs ci-après semblaient contribuer de manière importante à l'obtention d'un degré de comparabilité élevé entre l'EP et l'EC :

- clarté des Nds/LdM concernant la mesure de l'évolution de la performance par rapport à l'évaluation PEFA précédente ;
- emploi de la même équipe d'évaluateurs pour l'EP et l'EC ;
- établissement d'une structure spécifique pour la participation des autorités publiques à l'évaluation ;
- organisation d'ateliers d'ouverture et de clôture à l'intention des principales parties prenantes ;
- soumission des versions préliminaires des rapports au Secrétariat PEFA pour commentaire ;

15. Parmi les facteurs mentionnés ci-dessus, on ne note d'absence de corrélation avec le degré de comparabilité des évaluations que pour l'examen des versions préliminaires des rapports par le Secrétariat PEFA, et ce pour la simple raison que toutes les évaluations comparatives incluses dans l'analyse avaient été examinées par le Secrétariat. Les précédents rapports d'étape ont toutefois montré que l'examen du Secrétariat contribue significativement à améliorer le taux de respect du Cadre et, par conséquent, la mesure dans laquelle les notes attribuées aux indicateurs sont dûment justifiées par des éléments concrets. L'examen par le Secrétariat influe donc sur la robustesse de la comparaison des notes des évaluations préalables avec celles des évaluations comparatives.

16. L'analyse du degré de comparabilité de chacune des 74 composantes des indicateurs effectuée sur la base de toutes les évaluations comparatives montre que, en moyenne, 82 % des paires d'évaluations sont comparables lorsque l'on exclut les composantes ou indicateurs « non notés » (70 % dans le cas contraire). Le degré de comparabilité est supérieur à 80 % pour 51 des 74 composantes du Cadre. Le degré de comparabilité le plus faible (61%) a été constaté pour l'indicateur PI-7 (i) « *Importance des dépenses non rapportées de l'administration centrale* », pour lequel il est notoirement difficile de réunir des données factuelles. Lorsque les éléments « non notés » sont inclus dans les séries de données, les degrés de comparabilité diminuent et le classement des composantes change. Cette analyse peut aider à identifier les composantes des indicateurs pour lesquelles il conviendrait de présenter des précisions et de fournir des directives aux évaluateurs, notamment dans le cadre de cours de formation.

Tendances de la performance de la GFP

17. Sur l'ensemble des composantes des indicateurs auxquelles une note a été attribuée, 11 % ont conservé la note « A » qui est la plus haute pouvant être donnée, 21 % ont amélioré leur note, 10 % ont conservé la note « D », qui est la plus faible possible, 9 % ont affiché une dégradation de leur note, 21 % ont conservé la note « B » ou « C », et 30 %¹ n'ont pas pu faire l'objet d'une évaluation acceptable de l'évolution de la situation. La structure des changements a été analysée pour *tous les indicateurs ou composantes pour lesquels il a été possible de procéder à la comparaison des résultats* entre deux évaluations successives. On considère que ces variations représentent, pour l'essentiel, l'évolution réelle de la performance, bien qu'on ne puisse exclure que, dans certains cas, les nouvelles notes reposent sur de meilleures informations ou reflètent des jugements ou des interprétations différentes, qui n'ont pu être dépistées dans le cadre du présent exercice.

¹ Le total dépasse 100 % parce que les pourcentages ont été arrondis. Les 30% comprennent les « non notés ».

18. L'évolution tendancielle des performances a été analysée au moyen d'une méthodologie qui classe les indicateurs ou les composantes dans des catégories différentes selon qu'ils se rapportent à des aspects *institutionnels* de la GFP pour lesquels des progrès peuvent résulter de l'adoption d'une loi, d'une réglementation ou d'une technique nouvelle, ou d'une action ciblée sur un petit nombre d'organismes ou menée à un stade précoce du cycle budgétaire, ou bien à des aspects *fonctionnels* de la GFP pour lesquels l'obtention de progrès passe nécessairement par la mise en application d'une nouvelle loi ou réglementation, ou la coordination des travaux de nombreux organismes, ou encore la poursuite d'une action à une phase plus tardive du cycle budgétaire ; en d'autres termes, cette méthodologie classe les composantes des indicateurs en fonction des caractéristiques des éléments de la GFP auxquels elles se rapportent pour établir une distinction en termes de réformes *de jure/de facto*, d'action en amont/en aval et de contrôle concentré/déconcentré.

19. L'analyse montre que les éléments institutionnels sont davantage susceptibles d'améliorer leur note ou de conserver la note maximale, tandis que les éléments fonctionnels sont davantage susceptibles d'enregistrer une détérioration de leur note ou de conserver la note minimale. Quelle que soit la catégorie considérée (institutionnelle ou fonctionnelle), la proportion des éléments affichant des notes maximales ou en hausse est supérieure à celle des éléments affichant des notes minimales ou en diminution, mais les écarts sont plus grands dans la catégorie des éléments institutionnels que dans la catégorie des éléments fonctionnels. Plus précisément :

- Les indicateurs décrivant le degré de concentration des intervenants présentent des améliorations de la performance bien plus marquées que les indicateurs pour lesquels le contrôle est déconcentré. Dans ce dernier cas, il n'est guère possible d'identifier quelque amélioration générale que ce soit.
- Les indicateurs qui couvrent des actions « en amont » et « en aval » du cycle budgétaire affichent des résultats similaires.
- Les indicateurs correspondant à des réformes *de jure* présentent des améliorations un peu plus marquées que ceux qui couvrent les réformes *de facto*.
- Les écarts entre les notes des éléments institutionnels et celles des éléments fonctionnels sont plus grands pour les notes minimales ou décroissantes que pour les notes maximales ou en hausse.

20. Sept évaluations comparatives réalisées moins de 30 mois après l'évaluation préalable ont été analysées dans le but de déterminer si la mesure de la performance présente, dans leur cas, des caractéristiques particulières.). Alors que l'on s'attendait à ce que les progrès accomplis sur des périodes inférieures trois ans ne soient guère perceptibles, il s'avère que le nombre de notes en amélioration est en fait plus élevé pour les évaluations effectuées à des intervalles plus courts, tandis que le nombre de notes non comparables est plus faible. Cela pourrait s'expliquer par le fait que la poursuite d'évaluations comparatives relativement rapidement après l'évaluation préalable a été principalement motivée par le fait que la réalisation d'une EC était une condition nécessaire à l'obtention de l'appui des bailleurs, ce qui a pu inciter les autorités à présenter le système de GFP de la manière la plus favorable possible. Il est aussi possible que les autorités et les bailleurs aient été incités à procéder à une évaluation répétée précoce si les parties prenantes étaient convaincues que l'évaluation répétée indiquerait une amélioration significative des systèmes de GFP.

Recommandations

21. Sur la base de ces observations, un certain nombre de recommandations sont présentées, dont certaines figurent déjà dans de précédents rapports d'étape. *Le Secrétariat PEFA* devrait prendre les mesures suivantes :

- Revoir les parties des documents existants consacrées au processus d'évaluation (« Liste de contrôle des LdM », « Bonnes pratiques pour l'application du Cadre de mesure de la performance de la GFP », « Note d'information pour les évaluations répétées ») afin de faire ressortir la nécessité : i) d'inclure dans les NdS/LdM une référence spécifique aux notes d'orientation du Secrétariat ; ii) de prévoir des délais et ressources supplémentaires pour l'analyse des changements ; iii) de soumettre les NdS/LdM au Secrétariat pour commentaire ; et iv) de veiller à ce que des informations détaillées sur l'évaluation préalable soient communiquées aux évaluateurs chargés de l'évaluation comparative, en recourant à certains des mêmes évaluateurs (ou, au moins, en sollicitant la collaboration du chef de l'équipe de l'évaluation préalable), et en transmettant aux évaluateurs de l'EC les commentaires des parties prenantes et du Secrétariat PEFA sur l'évaluation préalable.
- Revoir les matériels de formation existants pour mettre en relief les points précédents.
- Examiner les composantes des indicateurs pour lesquels il est rare de pouvoir procéder à des comparaisons pour déterminer s'il peut être nécessaire d'apporter des précisions au cadre ou de communiquer des directives aux évaluateurs, et s'il convient d'envisager de modifier les normes minimales de notation des composantes.

22. *Les institutions chef de file* devraient veiller au respect d'un certain nombre de bonnes pratiques :

- Les NdS/LdM doivent préciser aussi clairement que possible ce qui est attendu de l'évaluation répétée, comme le recommande la « Note d'information pour les évaluations répétées » du Secrétariat PEFA.
- Les NdS/LdM doivent indiquer de façon suffisamment détaillée aux évaluateurs comment présenter les comparaisons effectuées avec l'évaluation préalable dans le rapport de l'évaluation comparative et prévoir le temps nécessaire pour permettre aux évaluateurs de vérifier la base sur laquelle les notes précédentes ont été attribuées.
- Les NdS/LdM doivent demander à l'équipe d'évaluation d'enregistrer toutes les informations pertinentes (sur un CD par exemple) pour qu'elles soient aisément accessibles et compréhensibles par d'autres experts.
- Les NdS/LdM doivent être inclus dans une annexe de la version préliminaire/finale du rapport d'évaluation.
- Les NdS/LdM doivent être soumis à une procédure d'assurance de la qualité (examen par les pairs) par le Secrétariat PEFA pour garantir le respect en bonne et due forme des points précédents.

- Il importe d’instaurer une procédure pour assurer la communication à l’équipe d’évaluateurs chargée de l’EC de tous les renseignements et documents pertinents se rapportant aux évaluations préalables, notamment le rapport définitif et les commentaires des parties prenantes et du Secrétariat.
- Les institutions chef de file doivent faciliter la prise de contacts entre le chef de l’équipe de l’évaluation préalable et de l’évaluation en cours, même si cela implique des délais supplémentaires et entraîne des coûts additionnels.
- En cas de désaccord avec les commentaires du Secrétariat relatifs aux NdS/LdM et aux évaluations, il importe de communiquer au Secrétariat les raisons de ces désaccords.

23. Lorsqu’ils procèdent à une évaluation répétée, *les évaluateurs* doivent :

- Suivre les directives formulées par le Secrétariat PEFA dans la « Note d’information pour les évaluations répétées ».
- Demander au responsable de l’évaluation (institution chef de file) toutes informations pertinentes relatives à l’évaluation précédente (versions préliminaire et finale du rapport et commentaires formulés dans le cadre de la procédure d’examen de la qualité).
- Demander à l’institution chef de file d’établir des contacts avec l’équipe chargée de l’évaluation préalable.

Poursuite des travaux

24. Les observations et recommandations précédentes donnent une idée de la contribution que pourrait avoir l’expansion de la base de données sur les évaluations PEFA répétées à l’analyse de l’évolution de la performance de la GFP. Le rapport présente de nombreuses suggestions sur la manière dont la *poursuite des travaux* pourrait renforcer les conclusions ou permettre d’explorer plus en détail les questions de fond. Si le Secrétariat PEFA entend poursuivre certains travaux, en particulier ceux qui portent sur la nécessité de développer le Cadre PEFA et de fournir des directives pour son application, de nombreuses autres possibilités s’offrent à la communauté des chercheurs, en particulier en ce qui concerne l’évolution de la performance des systèmes de GFP.

Chapitre 1 – Introduction et méthodologie

25. Ce rapport d'étape est le quatrième préparé par le Secrétariat. La mesure des progrès accomplis dans le temps est l'un des objectifs premiers du Cadre PEFA. Jusqu'ici, la plupart des évaluations entreprises ont été des évaluations de référence, mais le nombre d'évaluations répétées commence à être significatif parce que les évaluations de référence ont été réalisées pour presque tous les pays dans certaines régions et que les évaluations initiales remontent maintenant à au moins trois ans.

26. Une évaluation répétée a notamment pour objectif de mesurer l'évolution de la performance depuis l'évaluation précédente. Elle examine les changements spécifiques survenus dans la performance du système en considérant ce qui s'est modifié et dans quelle mesure. Les parties prenantes peuvent compter sur le fait que, si le Cadre est appliqué de manière systématique à différentes reprises, les évaluations qui se succèdent fourniront des indications sur l'évolution de la performance de la GFP d'un pays. Le rapport sur la performance de la GFP fera en outre état des actions entreprises par les autorités publiques pour réformer le système de GFP en décrivant les mesures de réforme qu'elles ont récemment adopté ou mises en application, même si elles n'ont pas encore eu d'impact sur ses performances. Il est immédiatement évident que, pour que deux évaluations séparées par un certain laps de temps soient comparables, il faut que la méthode suivie ait été appliquée de manière cohérente. Si l'on part du principe que l'évaluation initiale est fondée sur une application rigoureuse de la méthodologie, les parties prenantes devraient, dans l'idéal, pouvoir compter sur une mesure exacte des progrès accomplis par la suite.

27. À présent qu'un nombre considérable d'évaluations répétées sont disponibles, le principal objet du *Rapport d'étape 2010* consiste à déterminer si le cadre PEFA permet de mesurer de manière fiable l'évolution de la performance dans le temps². Les résultats de l'analyse contribueront dans une mesure considérable à déterminer s'il importe de fournir de plus amples orientations techniques et directives sur les processus. L'analyse permettra également de donner des directives aux fins de la comparaison des notes attribuées aux indicateurs PEFA pour différentes années et apportera des éléments d'information importants pour l'évaluation externe de l'ensemble du Programme PEFA.

28. Le rapport vise à répondre à trois grandes questions : i) quelles sont la fréquence et les raisons d'être des évaluations répétées ? ii) Le Cadre permet-il de mesurer de manière effective l'évolution des performances, et serait-il possible d'améliorer la fiabilité et la validité des mesures ? iii) Quelles tendances de la performance de la GFP les évaluations répétées font-elles apparaître ?

29. Aux fins de la préparation de ce rapport, l'équipe a adopté les définitions suivantes:

i) Une « *évaluation préalable* » (EP) est une évaluation qui mesure les systèmes de GFP du pays et la performance de ses processus à un certain moment, conformément à la méthodologie PEFA. L'évaluation peut avoir pour but d'établir un niveau de référence à

² L'étude n'a pas entrepris d'analyser la totalité du rapport d'évaluation car ce n'était pas son but ; elle portait sur des thèmes spécifiques de la Synthèse de l'évaluation et des sections 1 et 2, et sur la qualité de la comparabilité des indicateurs/composantes de la section 3.

partir duquel mesurer les progrès de la performance de la GFP mesurée au niveau des indicateurs ou des composantes³ au cours des années suivantes.

ii) Une « *évaluation comparative* » (EC) est une évaluation PEFA répétée qui fait référence à l'évaluation préalable, compare les notes des deux évaluations, et tente de rendre compte des différences et des évolutions de la performance de la GFP.

30. En s'appuyant sur ces définitions, le rapport a pris en considération 33 évaluations répétées (annexe A), considérées être des « *évaluations comparatives* » parce qu'elles visaient à mesurer les changements survenus sur la durée. Douze évaluations répétées qui ne répondaient pas aux critères requis pour être considérées comme des EC (voir à l'annexe B une brève explication de chaque cas) n'ont pas été prises en compte.

31. Le chapitre 2 analyse la fréquence des évaluations répétées et ses déterminants. Le chapitre 3 évalue dans quelle mesure le Cadre permet effectivement de mesurer les changements intervenus. Le chapitre 4 résume les tendances de la performance de la GFP telles qu'elles ressortent des évaluations répétées et le chapitre 5 présente les recommandations.

32. La méthodologie utilisée pour les chapitres 2, 3 et 4 se présente comme suit :

33. **Le chapitre 2** examine les raisons lesquelles les évaluations comparatives ont été effectuées, d'une part en analysant les motifs indiqués dans les rapports et, d'autre part, en demandant aux parties prenantes ayant financé ou entrepris des EC quelles autres raisons elles peuvent avoir eu. L'analyse a porté sur les versions initiales et finales des rapports, ainsi que sur les examens de suivi du Secrétariat et les commentaires de ce dernier sur la synthèse de l'évaluation, la section 1 et la section 2. Des entretiens ont été menés par téléphone et par courriel avec de nombreuses parties prenantes (représentants des institutions chef de file et chefs des équipes d'évaluation) ayant financé ou conduit des évaluations PEFA répétées. Environ 70 % d'entre elles ont accepté de participer aux entretiens. Il est arrivé que des parties prenantes expriment, durant les entretiens, des opinions différentes de celles contenues dans les rapports écrits, de sorte qu'il a fallu effectuer certains jugements pour évaluer la situation. Il est, par exemple, souvent difficile de déterminer si l'évaluation comparative a été conduite à la demande des bailleurs ou du pays concerné, et quelle en était la motivation principale. L'analyse a pris en compte les 33 EC qui avaient été finalisées ou achevées pour l'essentiel. Lorsque plusieurs EC ont été réalisées pour un pays, l'analyse a porté sur la plus récente, sauf dans le cas du Malawi où deux EC ont été analysées.

34. Dans le contexte des chapitres 3 et 4, l'EC analyse les progrès principalement en évaluant les indicateurs et leurs composantes auxquels une note a été attribuée. Les résultats des comparaisons des notes sont décrits par seize combinaisons possibles, qui sont présentées à l'annexe C. Si les quatre notes qui peuvent être attribuées (A, B, C ou D) aux composantes permettent de voir si leur performance s'est améliorée, ne s'est pas modifiée ou s'est dégradée, les combinaisons d'observations comparatives décrivent de manière plus détaillée l'évolution de la performance entre l'évaluation préalable et l'évaluation répétée.

³ L'analyse présentée dans ce rapport repose sur les notes attribuées aux composantes des indicateurs et sur celles des indicateurs eux-mêmes lorsqu'ils ne sont pas structurés en plusieurs composantes.

35. Les composantes auxquelles aucune note n'a été attribuée (désignées par l'expression « non notées ») sont aussi prises en compte dans certaines des analyses incluses dans ce rapport mais, comme il n'est pas possible de comparer deux évaluations lorsque l'une d'elles a reçu une mention « non noté », celles-ci font l'objet d'analyses distinctes⁴.

36. Dans la plus grande partie de ce rapport, les observations des évaluations comparatives sont présentées en pourcentage du nombre total des observations, y compris les éléments non notés. En effet, on compte que le nombre des observations non notées diminuera à mesure que les systèmes de GFP et les techniques d'évaluation PEFA s'amélioreront. C'est ce qui s'est produit dans le cas de l'échantillon, pour lequel le pourcentage moyen des observations non notées et de celles qui ne sont pas comparables pour d'autres raisons est passé de 29 % dans les EP à 20 % dans les EC.

37. Le **chapitre 3** examine les évaluations comparatives pour déterminer s'il est possible d'avoir confiance dans les résultats des comparaisons entre les notes des indicateurs et de leurs composantes et celles de l'évaluation précédente. Comme indiqué précédemment, ce sont les notes des composantes qui sont comparées, sauf si un indicateur n'a pas de composante, auquel cas c'est la note qui lui a été attribuée qui est utilisée. Il est donc nécessaire d'analyser les éléments concrets communiqués pour chaque indicateur et composante évalués pour vérifier qu'elles mesurent bien essentiellement la même chose. À cette fin ont été pris en compte les rapports des évaluations précédentes et des évaluations comparatives, les commentaires du Secrétariat et l'examen par ce dernier de la suite donnée à ses recommandations concernant les versions initiales et finales des évaluations préalables et des évaluations comparatives. Un tel examen a été réalisé lorsque cela n'avait pas déjà été fait. Cinq « *problèmes de suivi* » qui réduisent la mesure dans laquelle il est possible de se fier aux comparaisons ont été identifiés : a) la collecte de données sur des éléments nouveaux qui n'avaient pas été pris en compte dans l'évaluation précédente ; b) la modification de définitions ; c) les différences entre les interprétations ; d) les différences au niveau de l'échantillonnage ; et e) la réévaluation des notes de l'évaluation préalable dans le cadre de l'évaluation répétée.

38. Il a été tenu compte des Nds/LdM⁵ et des informations fournies par les rapports et les documents internes du Secrétariat sur les facteurs pratiques susceptibles d'affecter la qualité de la mesure des variations, à savoir l'emploi de la même équipe d'évaluateurs pour les évaluations, la participation des autorités à l'évaluation répétée, l'organisation d'ateliers d'ouverture et de clôture et la soumission des versions initiales des rapports au Secrétariat PEFA pour commentaires. Les Nds/LdM qui ont été préparées après la

⁴ L'absence de notation d'un indicateur ou d'une composante peut être due à de nombreuses raisons. Le Secrétariat PEFA a élaboré une méthodologie qui pour décrire les situations associées à l'absence de notes, qui sont, fondamentalement, au nombre de trois : Indicateur « non noté (NN), » lorsqu'on ne dispose pas d'informations suffisantes pour le noter, Indicateur « non applicable (NA), » lorsqu'un indicateur ou l'une de ses composantes n'est pas pertinent étant donné les conditions en vigueur dans le pays évalué, et Indicateur « non utilisé (NU), » lorsqu'il a été décidé de manière délibérée de ne pas noter l'indicateur ou l'une de ses composantes.

⁵ Le Secrétariat ne disposait que de moins de la moitié des versions finales des Nds/LdM lorsque la préparation du Rapport d'étape 2010 a commencé. Les Nds/LdM manquants ont été obtenus directement auprès des parties prenantes, tirés des versions finales des rapports, et dans un cas particulier communiqué par un consultant. Dans deux cas, aucune Nds/LdM n'avait été préparée.

publication par le Secrétariat PEFA de la note d'orientation intitulée « *When undertaking RA* » en février 2010, ainsi que l'évaluation répétée pertinente, ont également été prises en considération.

39. L'analyse présentée dans ce chapitre se fonde sur ce qui précède pour déterminer dans quelle mesure le Cadre facilite la mesure des changements à certains égards, et pour quelles raisons il ne les facilite pas à d'autres égards. Elle repose sur 33 évaluations comparatives dont 22 avaient été finalisées (rapport définitif) et 11 étaient achevées pour l'essentiel (la version initiale du rapport était achevée).

40. Lorsqu'il est possible d'avoir confiance dans les résultats des comparaisons entre les notes des indicateurs et de leurs composantes et leurs notes antérieures, le chapitre 4 passe en revue la nature et l'ampleur des changements ainsi que les indicateurs concernés. Il analyse aussi la structure des changements par catégorie d'indicateurs, caractéristiques nationales et régions en suivant une méthodologie (Andrews, 2009 ; Porter *et al.*, 2010) qui regroupe les indicateurs ou leurs composantes en fonction de trois critères : réformes *de jure/de facto*, mesures en amont/aval et contrôle concentré/déconcentré⁶.

- Le premier établit une distinction entre les composantes PEFA pour lesquelles une note d'au moins C pourrait résulter de la promulgation d'une nouvelle loi ou l'annonce d'une nouvelle, avant même leur mise en application (réforme **de jure**) et les composantes dont les notes sont fonction de la mise en œuvre effective des mesures annoncées ou du moins de la poursuite d'un effort significatif (réforme **de facto**).
- Le deuxième établit une distinction entre les composantes PEFA comme la budgétisation stratégique (prévisions pluriannuelles, planification stratégique, planification des investissements, planification de la dette), la préparation du budget annuel, l'analyse du budget annuel par le pouvoir législatif et la structure des documents budgétaires officiels (actions en **amont**), d'une part, et la gestion des ressources (notamment la gestion des entrées et sorties de trésorerie, la passation des marchés publics, la gestion de la masse salariale) ; le contrôle interne, l'audit et le suivi internes ; la comptabilité et la communication financière ; l'audit externe ; et l'analyse des rapports d'audit par le pouvoir législatif d'autre part (actions en **aval**).
- Le troisième établit une distinction entre les éléments PEFA qui sont placés sous le contrôle d'organes exécutifs centraux comme le ministère des Finances (contrôle **concentré**) et ceux qui dépendent de plusieurs ministères ou autorités infranationales (contrôle **déconcentré**).

41. L'évolution dans le temps est jugée principalement sur la base de sa direction, et non par le nombre de degrés franchis vers le haut ou le bas de l'échelle de notation. Les changements observés sont agrégés au niveau de toutes les composantes des indicateurs et de tous les pays, pour donner le même poids à toutes les composantes et à tous les

⁶ Cette méthodologie a été utilisée pour certains projets de recherche qui ont été largement diffusés. Bien que le Secrétariat PEFA n'approuve pas totalement cette classification pour tous les indicateurs, la méthodologie a été utilisée ici, telle quelle, car son adaptation exigerait beaucoup de temps et la participation de ceux qui l'ont initialement conçue. On pourrait par contre envisager d'engager des travaux additionnels pour améliorer la classification des indicateurs et examiner l'impact de ces améliorations sur les résultats de l'analyse.

pays. Le présent rapport n'a pas donné lieu à une analyse de la sensibilité des résultats à l'apport de modifications à cette méthode. Il serait utile d'examiner cette question de manière plus approfondie pour confirmer la robustesse des conclusions.

Chapitre 2 - Fréquence des ER et ses déterminants

2.1 Présentation générale

42. Cette section s'efforce de répondre aux questions suivantes :

- Quelle est la fréquence des évaluations répétées?
- Quels sont les raisons pour lesquelles une évaluation est répétée très fréquemment ou bien après la période de 3 à 5 ans recommandée ?
- Pourquoi certains pays n'ont-ils pas conduit d'évaluations répétées dans les délais recommandés ?
- Quels autres raisons —outre la mesure de l'évolution de la performance – ont pu motiver le lancement d'une EC ?
- Ces raisons varient-ils en fonction des caractéristiques nationales ? Du bailleur fournissant le financement ?
- Existe-t-il une relation entre la conduite de l'EC dans les délais recommandés et certaines caractéristiques nationales identifiables ?

2.2 Observations

43. Entre juin 2005 et le 6 octobre 2010, quarante-cinq évaluations répétées ont été entreprises ; ces évaluations, telles que celles qui les ont précédé y compris l'évaluation initiale, ont été menées conformément au Cadre PEFA. Elles ont été réalisées au niveau de 41 administrations nationales et infranationales dans 38 pays (trois évaluations répétées ont été effectuées au niveau infranational et une évaluation répétée a été menée au niveau national en Éthiopie). Sur ces 45 ER, 33 (indiquées à l'annexe A) remplissaient les critères requis pour être considérées comme une évaluation comparative, contrairement aux 12 autres (indiquées à l'annexe B).

44. Les EC ont été conduites entre un et quatre ans (en moyenne 33 mois) après l'EP⁷. Sept pays ont procédé à une EC moins de 30 mois (en moyenne 22 mois) après l'évaluation précédente⁸. Cet intervalle est plus court que le délai recommandé de 3 à 5 ans qui est jugé nécessaire pour que les réformes puissent produire des résultats mesurables. Il convient de se souvenir que l'intervalle entre les dates de l'EP et celles de l'EC ne reflète pas nécessairement l'intervalle séparant les séries de données utilisées. Par exemple, dans le cas du Kenya (où 30 mois se sont écoulés entre les missions), l'EP était basée sur les données de l'exercice 04/05 et l'EC sur les données de l'exercice 07/08 ; l'intervalle recommandé de trois ans entre les évaluations a donc été respecté.

45. Cela étant, les deux raisons les plus couramment citées pour expliquer la brièveté des intervalles entre deux évaluations PEFA sont la nécessité de faciliter le dialogue avec les bailleurs, et la nécessité de mesurer les progrès accomplis dans le cadre des réformes. Par exemple en Afghanistan, la Banque mondiale, qui finance environ 30 % du budget non militaire de l'État, avait besoin d'une EC pour obtenir les informations requises pour son cadre fiduciaire. En République dominicaine, l'évaluation a été justifiée notamment

⁷ Sur la base des dates des missions principales menées pour chaque évaluation. Dans les trois cas où il n'a pas été possible d'obtenir la date de la mission principale de l'EP, c'est la date du rapport qui a été utilisée.

⁸ Sur la base des dates des missions principales pour chaque évaluation.

par la volonté d'un dialogue plus systématique avec la communauté internationale, mais aussi, dans une mesure cruciale, par la nécessité pour les autorités nationales de suivre les progrès accomplis dans la gestion des ressources. Au Mozambique, l'EC devait fournir une base solide à l'évaluation des progrès dans le temps. En Tanzanie, l'un des buts déclarés était « ...de fournir une base au dialogue entre les autorités nationales et les bailleurs sur les futures réformes de la GFP. »

46. Au nombre des autres raisons invoquées figurent principalement la nécessité d'avoir une évaluation PEFA pour pouvoir concevoir un plan d'action ou un programme de réforme de la GFP ou en assurer le suivi, et le fait qu'une telle évaluation est une condition nécessaire à la poursuite des opérations de soutien budgétaire en cours ou prévues ou à la réalisation d'un futur projet de GFP. L'Union européenne a, par exemple, convaincu le Gouvernement du Malawi d'attendre pour préparer un plan d'action pour la réforme de la GFP que les résultats de la première EC soient disponibles ; ce plan d'action a alors contribué à l'augmentation considérable de l'appui au budget général. La deuxième EC du Malawi (qui est la dernière en date) a servi de base à un second Plan d'action de GFP qui, lui aussi, était une condition nécessaire à l'octroi par l'Union européenne d'un appui au budget général. L'importance accordée par les autorités nationales aux évaluations PEFA s'est accrue avec leur nombre : la première n'a guère retenu leur attention, la deuxième a été prise plus au sérieux parce qu'elle était liée à la fourniture d'un appui budgétaire général, la troisième a été examinée avec attention par les autorités et a contribué à la poursuite d'un utile dialogue entre ces dernières et les bailleurs finançant l'appui budgétaire général. Les parties prenantes ont reconnu que l'intervalle d'un an entre la première EP et la première EC était trop bref pour qu'il soit possible de détecter des progrès ; il semble que les autorités se soient attendues, dans une bien plus large mesure que les bailleurs, à ce que la première EC indique des améliorations significatives. Au Kosovo enfin, les modifications considérables apportées au cadre institutionnel et constitutionnel du pays ont conduit aussi bien les autorités que les bailleurs à demander qu'il soit procédé à une nouvelle évaluation prenant en compte ces transformations.

47. En ce qui concerne les 26 EC effectuées 30 mois⁹ ou davantage (37 mois en moyenne) après l'EP, les principales raisons mentionnés sont la mesure des progrès et la contribution à la conception ou au suivi d'un plan d'action ou d'un programme de réforme de la GFP, suivis par la facilitation du dialogue avec les bailleurs et les liens entre les évaluations et la poursuite du soutien budgétaire général en cours ou à venir. Par exemple, la Dominique a procédé à une EC parce que les autorités trouvaient l'EP déficiente et mal réalisée ; elles estimaient que l'EC leur fournirait une base solide à partir de laquelle elles pourraient élaborer un nouveau plan d'action pour la GFP. L'EC devait aussi permettre de déterminer si le pays pourrait continuer de bénéficier d'un appui budgétaire. En Ouganda, l'EC a contribué à la réalisation d'un examen annuel de la performance de la GFP, réduit les coûts de transaction des administrations publiques en facilitant le dialogue avec les bailleurs grâce à un pool d'informations ayant l'aval de tous, fourni une justification cruciale à la fourniture d'un appui budgétaire général, et permis de vérifier si l'appui apporté par un vaste fonds commun de soutien avait réellement fait progresser la GFP. Au Timor-Leste, une EC a été effectuée dans le cadre d'une évaluation de la transparence financière dans le contexte du *Code de bonnes pratiques de transparence des finances publiques* du FMI – c'est-à-dire un RONC budgétaire – afin d'évaluer la performance et de déterminer s'il y avait lieu de recadrer

⁹ Sur la base des dates des missions principales.

les efforts en matière de GFP. En République kirghize, l'EC a évalué les progrès des réformes de la GFP et a appuyé l'élaboration d'une conception plus stratégique et plus programmatique des réformes de la GFP dans une perspective à moyen terme. Aux Tonga, le ministère des Finances souhaitait connaître l'opinion, formulée de manière indépendante, de personnes jetant un regard neuf sur le pays, quant aux réformes qui seraient raisonnables, réalisables et susceptibles de procurer des avantages, et à celles qui ne rempliraient pas ces critères. Au Ghana, la nouvelle équipe gouvernementale désirait obtenir une description exacte de la situation de la GFP, et encourager les partenaires de développement à utiliser davantage les systèmes du pays.

48. Dans 12 cas, les évaluations réalisées à la suite d'évaluations préalables ne sont pas considérées comme des EC car elles ne font pas référence à l'évaluation précédente et n'analysent pas l'évolution de la performance de la GFP. Bien que les raisons soient différentes pour chaque évaluation, le tableau 1 montre que les EP ou les ER concernés souffrent de problèmes de qualité peut-être plus prononcés. En moyenne, dans ces 12 cas, le nombre de composantes non notées dans les EP était plus élevé et la proportion d'évaluations examinées par le Secrétariat PEFA plus faible¹⁰. Ces ER ont été réalisées en moyenne 30 mois¹¹ après l'évaluation préalable dans le but, principalement d'établir des données de référence, ce qui suggère un certain manque de confiance dans l'évaluation préalable. Par exemple, l'EP de la Bolivie n'a été réalisée qu'à moitié et ne repose que sur une partie du Cadre ; il a donc été nécessaire de réaliser une évaluation intégrale pour avoir une évaluation de référence complète. Dans le cas de la République centrafricaine, la Banque africaine de développement préparait une opération d'appui institutionnel à la GFP, et avait besoin d'une évaluation PEFA à jour en tant que base de référence. Le fait que l'EP n'ait pas été validée dans le cadre d'un atelier de clôture a été jugé préoccupant. De plus, pour pouvoir assurer un appui au budget général, la Banque africaine de développement et la Banque mondiale avaient besoin d'utiliser les indicateurs PEFA comme indicateurs de performance, et de disposer d'une évaluation à jour pour pouvoir finaliser le rapport de fin d'exécution. Dans le cas de Montserrat, l'EP de 2008 était jugée ne plus être à jour car elle reposait essentiellement sur une évaluation du risque fiduciaire menée par le DFID en 2006. Elle n'avait pas non plus été soumise au Secrétariat PEFA pour examen. Le DFID n'a donc pas pu l'utiliser en 2010 comme base de son évaluation du risque fiduciaire (comme l'exigeaient ses directives internes) et il est devenu nécessaire de réaliser une nouvelle EC.

Tableau 1: Différences entre les EC et les autres ER sur le plan de la couverture et de l'assurance de la qualité

Évaluations répétées	Nombre moyen de composantes non notées		Version initiale de l'évaluation examinée par le Secrétariat PEFA (%)	
	EP	ER	EP	ER
33 EC incluses	10	3	100 %	100 %
12 ER exclues	27	3	83 %	83 %

49. Il a été procédé à une analyse de corrélation pour déterminer si les raisons ou la fréquence des EC varient en fonction des caractéristiques nationales ou du bailleur apportant le financement, mais aucune tendance n'a pu être détectée. Il a été envisagé de

¹⁰ La liste complète des EP rejetées et les raisons de leur rejet sont indiquées à l'annexe B.

¹¹ Sur la base des dates indiquées en couverture des rapports.

procéder à une analyse de régression à plusieurs variables pour déterminer l'existence de facteurs éventuels. Il a toutefois été établi que, par suite de la petite taille de l'échantillon, du domaine de variation limité (intervalle entre les évaluations) de la variable dépendante, et de la nature subjective des variables explicatives pouvant être retenues (comme l'importance accordée à la présentation de progrès dans le cadre des réformes de la GFP), une telle analyse ne produirait probablement pas de résultats significatifs à ce stade.

50. Au total, 23 pays avaient achevé une évaluation de référence à la fin de juin 2006, c'est-à-dire durant l'année qui a suivi le lancement du Cadre. Près de cinq ans plus tard, 18 de ces pays (78 %) avaient entrepris une ER, dont six ne sont pas une évaluation comparative pour les raisons évoquées plus haut. Les cinq évaluations qui n'ont pas été suivies d'une évaluation répétées (voir annexe D) ont donné lieu à un examen visant à expliquer pourquoi une ER n'avait pas été entreprise dans l'intervalle recommandé. L'EP de Fidji a été finalisée en juin 2005. Aucune évaluation répétée n'a été lancée pour des motifs ayant trait aussi bien à la demande (les autorités sont réputées ne pas avoir été satisfaites de l'évaluation préliminaire) qu'à l'offre (les bailleurs ont réduit leur assistance après la prise du pouvoir par les militaires en 2006). Au Bangladesh, où un temps très long s'est écoulé avant la réalisation d'une nouvelle évaluation, l'adhésion des autorités nationales à l'évaluation PEFA menée en 2005 était très limitée, de sorte qu'il avait fallu employer une méthodologie modifiée. Cela a eu pour effet de remettre en question la validité de l'évaluation; les autorités n'ont pas jugé crédibles les notes attribuées et le rapport n'a pas été publié, à l'exception d'un tableau récapitulatif présenté en tant qu'annexe à la Stratégie d'aide de la Banque mondiale. Depuis lors, des diagnostics de la GFP portant une partie des éléments d'une analyse PEFA ont été effectués (notamment, Gouvernement du Bangladesh et DFID, 2007 ; Banque mondiale, 2006, 2007, 2009 ; DFID, 2009), et une évaluation comparative vient d'être entreprise. Dans le cas de la Syrie, une évaluation PEFA partielle a été incluse dans un rapport d'assistance technique du FMI. Aucun accord formel portant sur l'utilisation de la méthodologie PEFA n'avait été conclu avec les autorités syriennes, qui ne connaissaient pas cette dernière. L'absence d'adhésion des autorités a aussi manifestement posé un problème en République du Congo (Brazzaville). En Ouganda, il est prévu de réaliser une évaluation répétée au niveau des administrations locales vers la fin de l'année 2011 ; l'absence d'appropriation par les autorités n'explique peut-être pas le retard avec lequel cette évaluation sera entreprise puisque, dans l'intervalle, plusieurs évaluations répétées ont été menées au niveau national. En règle générale, lorsqu'une ER n'est pas réalisée dans le délai de cinq ans recommandé, il semble que ce soit par suite du manque d'adhésion des autorités à l'évaluation PEFA initiale. Plusieurs autres facteurs pourraient également avoir joué (le degré de dépendance envers l'aide, le climat politique local) mais l'on ne dispose pas d'un nombre de cas suffisant pour pouvoir justifier une analyse plus poussée.

51. En conclusion, la vaste majorité des évaluations PEFA sont suivies d'une ER dans l'intervalle recommandé de 3 à 5 ans (ou légèrement plus court). La plupart des 33 évaluations comparatives ont été effectuées au moins 30 mois après l'évaluation préliminaire. Les principales raisons d'être de la conduite de l'évaluation comparative étaient, par ordre d'importance, la volonté de mesurer les progrès accomplis et de contribuer à la conception ou au suivi d'un plan d'action ou d'un programme de réforme de la GFP, de faciliter le dialogue avec les bailleurs et d'établir des liens avec une opération d'appui au budget général en cours ou future. Sept évaluations comparatives ont été effectuées moins de 30 mois après l'évaluation préliminaire. Dans la plupart des

cas, c'est parce qu'il était nécessaire aux bailleurs et/ou aux autorités nationale de mesurer les progrès sans plus tarder. Aucune tendance n'a pu être détectée au niveau de la distribution des objectifs et de la fréquence des évaluations comparatives par rapport aux caractéristiques nationales ou au bailleur fournissant l'assistance. Les 12 pays qui ont procédé à des évaluations répétées qui ne sont pas considérées comme des évaluations comparatives avaient rejeté l'évaluation préalable et n'avaient pas analysé les variations de la performance de la GFP. Le manque de confiance dans la qualité de l'EP a pu jouer un rôle déterminant à cet égard. Enfin, dans cinq cas, aucune évaluation répétée n'avait encore été conduite 5 à 6 ans après l'EP. Le manque d'adhésion des autorités au processus PEFA paraît avoir été un facteur important dans ces cinq cas.

Chapitre 3 – Efficacité de la mesure de l'évolution de la performance

3.1 Aperçu

52. Cette section examine les questions suivantes :
- Quel est le degré de fiabilité de la mesure des changements ?
 - Quels sont les indicateurs et leurs composantes dont les notes peuvent être comparées à celles de l'évaluation précédente pour déterminer les variations intervenues sur la base de faits concrets ?
 - Pour quelles raisons les notes de certains indicateurs et composantes ne peuvent-elles être comparées de manière fiable avec les notes des évaluations précédentes (par ex : qualité des éléments concrets, explications et doutes relatifs aux notes précédentes ; clarté des instructions des LdM concernant la mesure des progrès dans le temps) ?
 - Quels « facteurs inhérents aux processus » influent sur la qualité de la mesure des changements ?

3.2 Constatations

Contexte

53. Le présent rapport n'a examiné que 33 évaluations répétées, considérées être des « évaluations comparatives » (EC) pour analyser la performance dans le temps. Au 6 octobre 2010, 22 évaluations comparatives avaient été finalisées (version finale du rapport) et les 11 autres étaient achevées pour l'essentiel (version initiale du rapport). Sur les 22 EC finalisées, 10 ont été rendues publiques¹². Les 33 évaluations comparatives couvraient 29 pays, dont 14 se trouvent en Afrique (AFR), six en Amérique latine (LAC), quatre dans la région Asie de l'Est et Pacifique (EAP), quatre en Europe de l'Est et Asie centrale (ECA) et une en Asie du Sud. Aucune évaluation comparative n'a été entreprise dans les pays de la région Moyen-Orient et Afrique du Nord (MENA) bien que trois années se soient écoulées dans certains pays (par exemple en Syrie) depuis la dernière évaluation PEFA.

54. L'analyse des progrès repose principalement sur une évaluation des indicateurs ou composantes auxquels une note a été attribuée¹³. La comparabilité des notes est étudiée à la fois au moyen d'une « analyse verticale » (les notes attribuées à chaque composante de chaque indicateur dans une évaluation comparative sont comparées à celles données dans l'évaluation préliminaire) et d'une « analyse horizontale » (les notes attribuées à une composante d'indicateur déterminée sont comparées à celles attribuées à la même composante dans toutes les évaluations comparatives).

¹² Les hyperliens intégrés dans le site du PEFA à l'adresse www.pefa.org permettent d'accéder aux rapports sur les pays suivants : Ghana, Guinée-Bissau, Kenya, Kosovo, Moldova, Ouganda, République kirghize, Timor-Leste, Trinité-et-Tobago et Tongas. Une évaluation comparative concernant la République dominicaine a de surcroît été publiée en janvier 2011.

¹³ Seules les notes des composantes sont analysées. Certains indicateurs n'ayant qu'une seule composante, c'est la note de l'indicateur qui a été utilisée.

55. Lorsque la mention « non noté » est attribuée à une composante dans l'une ou dans les deux évaluations, il devient impossible de comparer les indicateurs correspondants. Comme indiqué au Chapitre 1, les indicateurs ou composantes qui n'ont pas été notés sont également pris en compte dans l'analyse présentée ici mais, comme il est impossible de comparer les résultats de l'évaluation préliminaire et de l'évaluation comparative, ils sont examinés séparément. Ce point est important lorsque l'on analyse le degré de comparabilité des données. Les tableaux 2, 3 et 3a illustrent l'impact des mentions « non noté » dans l'analyse verticale et dans l'analyse horizontale. Lorsque les éléments « non notés » sont inclus dans la série de données, le niveau de comparabilité des 74 composantes est de 70 %. Lorsqu'ils en sont exclus, le niveau de comparabilité atteint 82 %. Cela signifie que 30 % de l'absence de comparabilité est due aux éléments « non notés » qui sont clairement indiqués dans les rapports d'évaluation. Les écarts observés sont similaires en ce qui concerne la comparabilité des différentes évaluations comparatives.

Tableau 2. Comparabilité des évaluations comparatives

<i>Fourchette de comparabilité des notes</i>	<i>Fourchette 80-100 %</i>	<i>Fourchette 50-79 %</i>	<i>Fourchette ≤ 50 %</i>	<i>Total EC</i>
Répartition des EC, éléments « non notés » inclus, dans chaque fourchette	15	11	7	33
Répartition des EC, éléments « non notés » exclus, dans chaque fourchette	25	5	3	33

Tableau 3. Comparaison des indicateurs/composantes

Pourcentage des indicateurs/composantes dont la comparaison peut être jugée fiable, lorsque les éléments « non notés » sont inclus	70 %
Pourcentage des indicateurs/composantes dont la comparaison peut être jugée fiable, lorsque les éléments « non notés » sont exclus	82 %

Tableau 3a. Comparaison des composantes des indicateurs (ventilation par fourchette de comparabilité et par nombre d'indicateurs)

<i>Fourchette de comparabilité des notes</i>	<i>Fourchette 80-100%</i>	<i>Fourchette 60-79%</i>	<i>Fourchette 50-59%</i>	<i>Fourchette ≤ 49%</i>	<i>Total des indicateurs et composantes</i>
Nombre d'indicateurs lorsque les éléments « non notés » sont inclus*	19	43	7	5	74
Nombre d'indicateurs lorsque les éléments « non notés » sont exclus*	51	23	0	0	74

* Les trois composantes de l'indicateur relatif à la prévisibilité des transferts des échelons supérieurs des administrations publiques ont été exclues de l'analyse. Voir le paragraphe 53.

Comparabilité au niveau des évaluations comparatives (analyse verticale)¹⁴

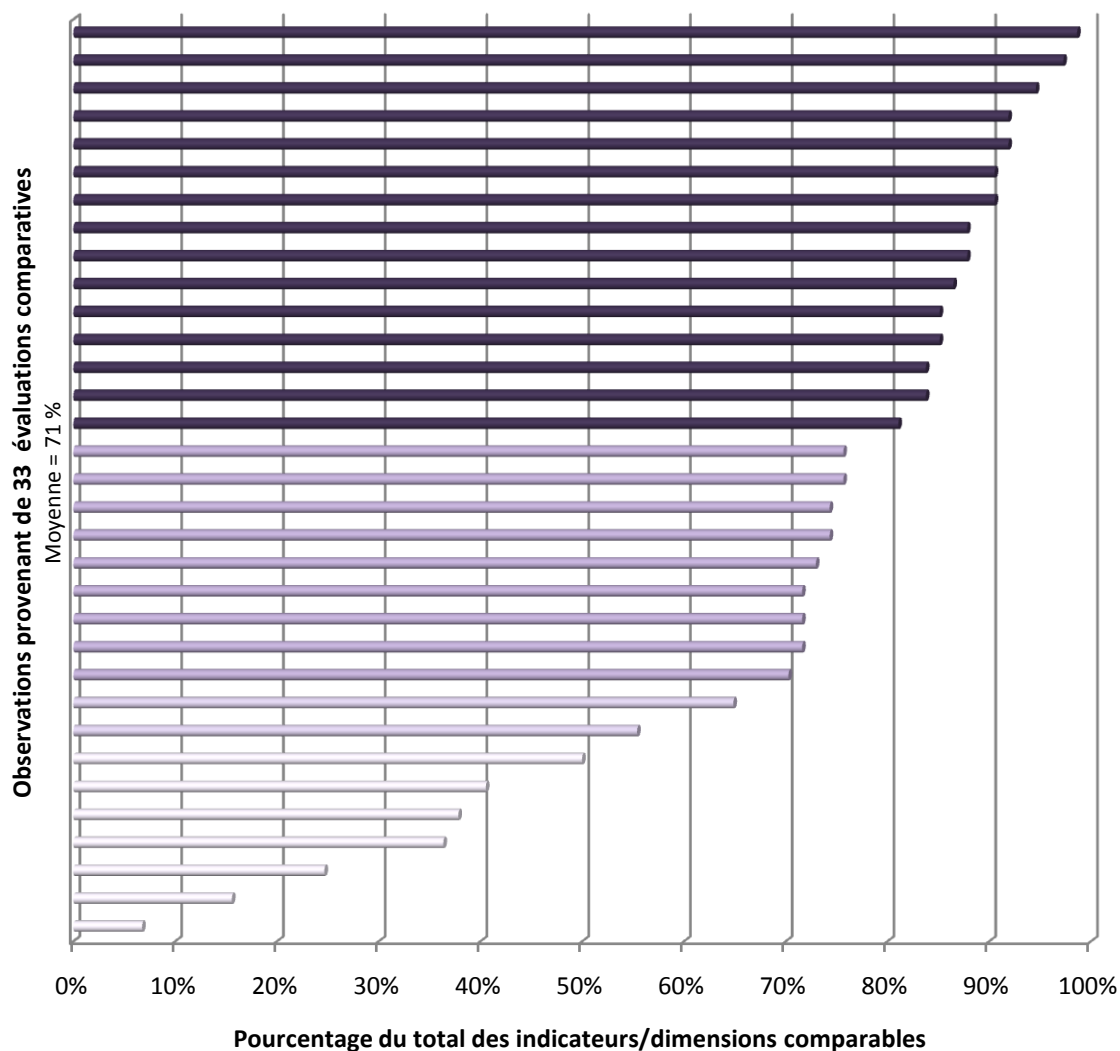
56. Sur l'ensemble des 33 évaluations comparatives, il s'avère qu'il est possible d'effectuer des comparaisons fiables entre 80 à 100 % des notes des indicateurs/composantes et les notes attribuées lors des évaluations préalables dans quinze cas (voir Graphique 1)¹⁵ et que le degré de comparabilité est compris entre 70 et 80 % dans le cas de neuf autres évaluations. Ces 24 évaluations comparatives représentent, conjointement, 73 % de toutes les EC prises en considération dans ce rapport.

57. Sept évaluations comparatives sont atypiques en ce sens que le pourcentage d'indicateurs ou de composantes dont les notes peuvent donner lieu à des comparaisons fiables à celles des évaluations préalables est anormalement bas (moins de 50 %), principalement parce que l'un des éléments avait reçu la mention « non noté » dans au moins l'une des évaluations.

¹⁴ L'annexe E indique le pourcentage d'indicateurs et de composantes qui peuvent faire l'objet de comparaisons fiables au niveau des EC lorsque les éléments « non notés » sont exclus.

¹⁵ Deux évaluations comparatives effectuées après la publication, en février 2010, des directives à suivre pour les évaluations répétées sont incluses dans ces 15 EC.

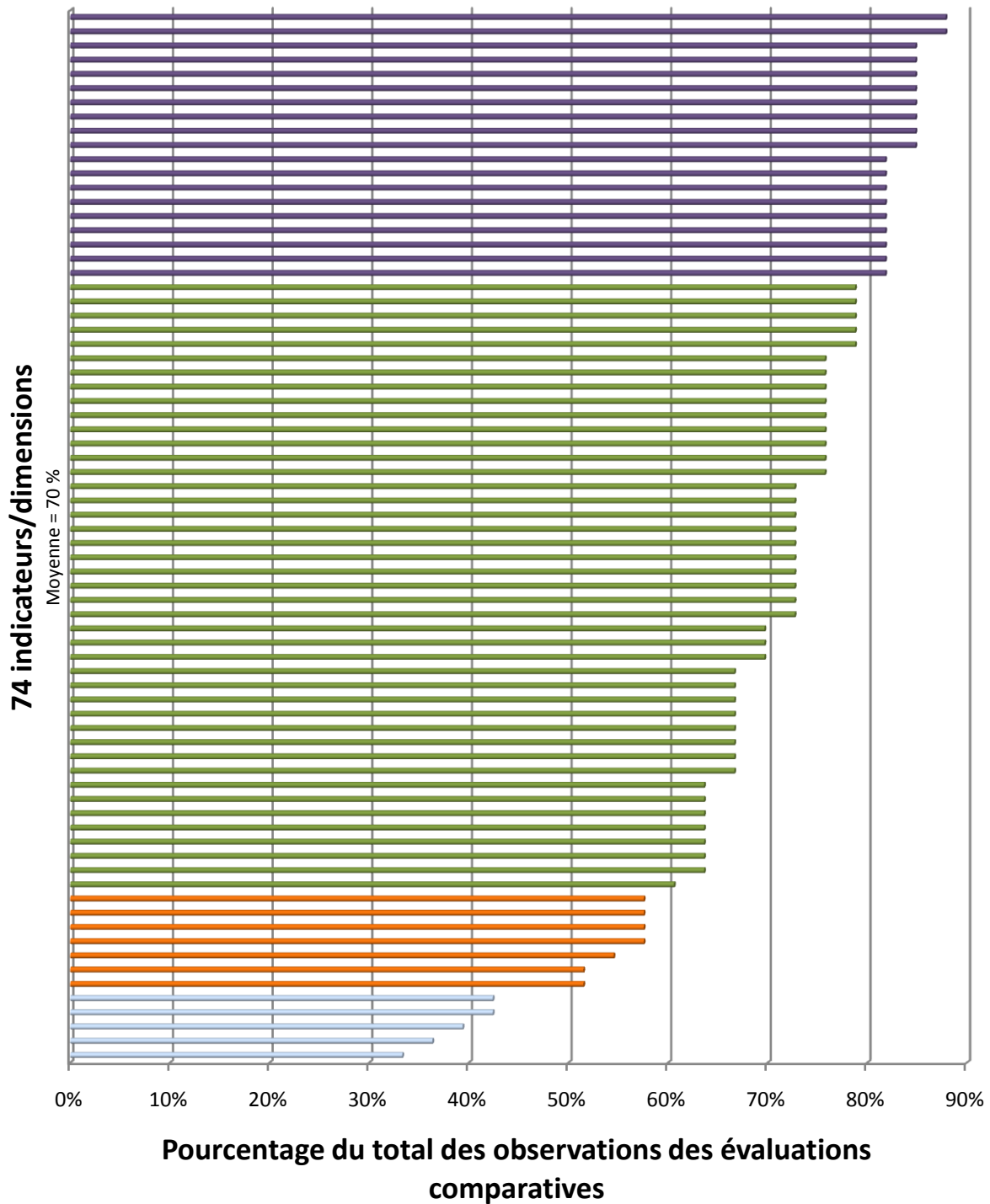
Graphique 1 : Pourcentage d'indicateurs ou de composantes dont les notes peuvent faire l'objet de comparaisons fiables entre évaluations comparatives (éléments « non notés » inclus)



Comparabilité des indicateurs ou composantes (analyse horizontale)

58. Les données provenant des 33 EC montrent que, en moyenne, les notes de 70 % des indicateurs ou des composantes sont comparables à celles de l'évaluation préalable. Il n'est pas possible d'effectuer des comparaisons fiables dans le cas d'environ 30 % des indicateurs ou des composantes en raison de l'absence d'éléments concrets, parce qu'ils ont reçu la mention « non noté » dans au moins l'une des deux évaluations, parce que les notes attribuées dans l'évaluation préalable ont été remises en question ou pour d'autres raisons encore. Étant donné que la plupart des évaluations préliminaires ont été réalisées au cours des deux années qui ont suivi la mise en œuvre du Cadre, époque à laquelle la qualité des évaluations était sensiblement inférieure à ce qu'elle est aujourd'hui, on peut considérer que ces pourcentages sont satisfaisants. On peut s'attendre à ce que, à l'avenir, les taux de comparabilité obtenus dans le cadre d'exercices similaires augmentent.

Graphique 2 : Pourcentage des notes attribuées aux différentes composantes d'indicateurs (éléments « non notés » inclus) qui peuvent faire l'objet de comparaisons fiables au niveau des évaluations comparatives



Légende¹⁶:

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

Indicateurs ou composantes pour lesquels le pourcentage de comparaisons fiables est élevé/faible

59. Le pourcentage de comparabilité des notes avec celles de l'évaluation préalable est compris entre 80 et 100 % pour 19 indicateurs ou composantes. En revanche, il est inférieur à 50 % pour cinq indicateurs ou composantes. L'indicateur de la prévisibilité des transferts des échelons supérieurs des administrations publiques n'a été noté que dans trois EC (aux fins de l'évaluation de l'appui au budget général). Cet indicateur n'existait pas lorsque l'évaluation de référence a été effectuée, si bien que toute comparaison aurait été impossible (0 % de comparabilité) ; il a donc été omis du graphique pour éviter de fausser les conclusions. Le graphique 2 indique les pourcentages des différents indicateurs/composantes (à l'exception des composantes de la prévisibilité des transferts des échelons supérieurs des administrations publiques) qui peuvent donner lieu à des comparaisons fiables sur l'ensemble des 33 évaluations comparatives.

60. Les cinq indicateurs ou composantes dans le cas desquels moins de 50 % des notes de l'évaluation comparative peuvent être comparées à celles de l'évaluation préalable sont :

- PI-4 (i), Stock et suivi des arriérés de paiement sur les dépenses
- PI-7 (i), Importance des opérations non rapportées de l'administration
- PI-15 (i), Efficacité du recouvrement des contributions fiscales et douanières
- D-1 (i) et (ii), Prévisibilité de l'appui budgétaire direct

61. La méthodologie PEFA exige des données quantifiables pour ces indicateurs ou composantes. Ce n'est pas le cas pour les 19 indicateurs/composantes affichant un taux de comparabilité très élevé à l'exception de PI-12 (iii) (qui requiert la détermination des coûts des stratégies sectorielles). Le problème rencontré pour les cinq indicateurs concernés pourrait tenir au manque de disponibilité des données (qui doivent être fournies par les autorités et par les partenaires de développement) et non pas nécessairement à la capacité de mesurer l'évolution de la performance à partir du Cadre. L'examen approfondi de la situation montre que le nombre de mentions « non noté » attribuées à ces cinq indicateurs ou à leurs composantes était considérable dans au moins l'une des évaluations, ce qui rendait toute comparaison impossible. Quatorze autres

¹⁶ Les indicateurs sont énumérés par ordre décroissant. Par exemple, l'indicateur PI-10 est le plus proche de 100 % tandis que PI-25 (ii) est le plus proche de 80 %.

indicateurs ou composantes ont reçu la mention « non noté » dans de 15 % des cas¹⁷. Si l'on part du principe qu'il était difficile de mesurer l'évolution de la performance de la GFP les premières années, on peut penser que les problèmes de comparabilité décroîtront à mesure que les autorités amélioreront leur capacité à produire les informations nécessaires et à mettre en œuvre des réformes. Il conviendra cependant de poursuivre l'analyse afin de déterminer si les difficultés posées par les comparaisons exigent : i) une modification des conditions de base à remplir pour chaque note pouvant être attribuée à une composante, ou ii) l'apport de meilleures « *Précisions* » pour l'application du Cadre, ou si une préparation insuffisante des missions PEFA peut avoir un impact sur le degré de comparabilité parce les données requises ne sont pas toujours disponibles.

62. Le graphique présenté à l'annexe F indique le taux de comparabilité des indicateurs lorsque les éléments « non notés » sont exclus : il montre que le taux de comparabilité se situe dans la fourchette 80-100 % pour trois-quarts des indicateurs ou composantes et dans la fourchette 60-80 % pour le quart restant ; aucun indicateur ou composante n'affiche de taux de comparabilité inférieur à 60 %. Comme dans le graphique 2, il est fait abstraction de l'indicateur de la prévisibilité des transferts des échelons supérieurs des administrations publiques qui n'est inclus dans aucune évaluation préliminaire.

« *Problèmes de suivi* »

63. Cinq « *problèmes de suivi* » qui ont fait obstacle aux comparaisons ont été identifiés : a) la collecte de données sur des éléments nouveaux qui n'étaient pas disponibles ou n'avaient pas été pris en compte dans l'évaluation précédente ; b) la modification de définitions ; c) les différences entre les interprétations des critères du Cadre ; d) les différences au niveau de l'échantillonnage ; et e) la réévaluation des notes manifestement erronées de l'évaluation préalable dans le cadre de l'évaluation répétée. Il importe ici de souligner que cette méthode, qui vise à assurer la fiabilité des comparaisons, est extrêmement prudente. Par exemple, lorsque l'équipe s'est aperçue que les notes des évaluations préalables avaient été recalculées dans le cadre des évaluations comparatives dans le but d'identifier les changements intervenus, il a été décidé de ne pas déterminer la direction de l'évolution dans le temps parce que les données ne permettaient pas de procéder à des comparaisons fiables. Les directives relatives aux évaluations répétées (février 2010) stipulent que les erreurs de notation manifestes détectées dans les évaluations préalables ne doivent pas être rectifiées, mais signalées dans le cadre de l'examen des changements¹⁸.

64. L'examen des 33 évaluations comparatives, auxquelles sont associées 2 451 observations, a permis d'identifier 175 cas de problèmes de « suivi » dont : i) 24 % tenaient à la collecte de données sur des éléments nouveaux concernant l'évaluation préalable, (ii) 14 % à la modification de définitions, (iii) 45 % à des différences d'interprétation, (iv) 6 % à des différences au niveau de l'échantillonnage, et v) 11 % à la réévaluation, dans le cadre de l'évaluation comparative, des notes attribuées lors de l'évaluation préalable. Certains « *problèmes de suivi* » sont peut-être plus faciles à éviter que d'autres (notamment les différences d'interprétation et d'échantillonnage et la

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

¹⁸ Il importe de noter qu'il sera nécessaire de recalculer les notes de trois indicateurs dont les échelles de notation ont récemment été révisées. Les équipes d'évaluation devront réévaluer les notes attribuées à PI-2 (i), PI-3 et PI-19 (iii) dans l'évaluation préalable sur la base des nouveaux critères pour pouvoir déterminer la direction dans laquelle ces composantes ont évolué dans le temps.

réévaluation des notes attribuées dans l'évaluation préalable, par opposition à la collecte de données sur des éléments nouveaux concernant l'évaluation préalable et la modification de définitions). Les problèmes les plus faciles à éviter représentent 62 % du total ; toutefois si les « différences d'interprétation » sont un problème spécifique à l'évaluateur, les autres tiennent plutôt aux processus suivis. Il serait possible d'étudier de manière plus approfondie la question pour identifier les composantes qui sont plus exposées à des problèmes de suivi ainsi que les raisons pour lesquelles ces problèmes surviennent, et recommander les mesures appropriées. L'encadré 1 décrit quelques exemples de « problèmes de suivi ».

Problèmes liés aux processus susceptibles d'affecter la qualité de l'évaluation des changements

65. Les facteurs liés au processus d'évaluation tels que les NdS/LdM, l'emploi de la même équipe d'évaluation pour l'évaluation préalable et pour l'évaluation comparative, le degré d'adhésion des autorités à l'évaluation, l'organisation ou non d'ateliers d'ouverture et de clôture et les procédures d'examen du contrôle de qualité ont été examinés dans le but de déterminer ceux qui peuvent intervenir dans le processus et peuvent avoir un impact sur la qualité de l'évaluation des changements.

Encadré 1. Exemples de « problèmes de suivi »

Collecte de données sur des éléments nouveaux concernant l'évaluation préalable : l'évaluation en cours a permis d'obtenir des informations qui n'étaient pas disponibles lors de l'évaluation préalable.

Modifications de définitions : notamment pour les arriérés, arriérés intérieurs, fonds extrabudgétaires, classification des entreprises paraétatiques, les administrations infranationales par opposition aux organes déconcentrés de l'administration centrale.

Différences d'interprétation : par exemple pour PI-24 (ii) : l'évaluation comparative a établi que les rapports sur l'exécution du budget de l'exercice précédent étaient publiés de 4 à 7 semaines après la fin du trimestre (5,5 semaines en moyenne), et a attribué la note « B » à cette composante. L'évaluation préalable avait noté une performance identique mais avait attribué la note « C » à la composante parce que deux rapports avaient été soumis avec plus de six semaines de retard.

Différences d'échantillonnage : par exemple PI-21 : l'évaluation préalable avait été préparée sur la base d'informations sur la situation en vigueur dans les ministères de l'Agriculture, de la Justice et des Finances. L'évaluation comparative a utilisé les données des ministères de l'Éducation et de la Justice. Est-il possible de comparer directement les évaluations ?

L'évaluation comparative a réévalué les notes attribuées lors de l'évaluation préalable : par exemple lorsque : i) des notes ont été attribuées de manière erronée bien que les éléments concrets aient été tout à fait clairs ; ii) un « + » a été attribué à un indicateur n'ayant qu'une seule composante.

66. *Clarté des Notes de synthèse/des Lettres de mission en ce qui concerne l'évaluation des changements.* Les NdS/LdM de 31 évaluations comparatives ont été analysées ; en Éthiopie, l'évaluation menée au niveau de l'administration fédérale et les trois évaluations menées au niveau d'administrations infranationales étaient couvertes par les mêmes NdS/LdM qui, pour des raisons statistiques, ont toutefois été comptées comme quatre NdS/LdM. Deux évaluations comparatives n'ont pas donné lieu à la préparation de NdS/LdM.

67. Solliciter les commentaires du Secrétariat PEFA sur les projets de NdS/LdM est loin d'être une pratique courante. Au commencement de la préparation de ce rapport d'étape, moins de la moitié des NdS/LdM des évaluations considérées ici avaient été

communiqués au Secrétariat. Elles ont ultérieurement été obtenues auprès des parties prenantes, extraites des versions finales des rapports ou communiquées par des consultants. La version définitive n'est habituellement pas communiquée au Secrétariat après qu'il a fait connaître ses commentaires ; des NdS/LdM sont parfois transmises au Secrétariat trop tard pour qu'il puisse émettre des commentaires.

68. L'analyse révèle que 24 NdS/LdM (73 %) requièrent une évaluation de l'évolution de la performance depuis l'évaluation PEFA précédente. Cette évaluation est en général demandée par : 1) une *simple référence* à « la mesure des progrès de la GFP depuis la dernière évaluation », ou 2) une *description claire* de ce qui est attendu de l'évaluation comparative.

69. Si l'on considère les 15 évaluations comparatives dont 80 à 100 % des indicateurs ou composantes pouvaient faire l'objet de comparaisons fiables, les NdS/LdM correspondantes stipulaient, pour 14 d'entre elles, que les changements intervenus depuis l'évaluation précédente devaient être mesurés, tandis que la quinzième n'en faisait pas mention. Dans le cas des sept évaluations comparatives dont les indicateurs avaient un degré de comparabilité limité (moins de 50 %), les NdS/LdM de quatre d'entre elles demandaient que l'évolution de la performance soit mesurée (trois évaluations au niveau infranational étaient associées à une évaluation au niveau de l'administration centrale), deux n'en faisaient pas mention, et aucune NdS/LdM n'avait été préparée dans le dernier cas¹⁹.

70. Ceci confirme qu'il importe « *de faire en sorte que les NdS/LdM soient suffisamment détaillées et bien comprises par toutes les parties prenantes* » comme le recommande le Secrétariat PEFA dans la note d'information sur les *Bonnes pratiques pour l'exécution des évaluations répétées*.

71. Il ressort d'observations ponctuelles que les NdS/LdM des évaluations répétées devraient prévoir le temps nécessaire pour vérifier les bases sur lesquelles les notes précédentes ont été attribuées. Bien que cette tâche soit une étape implicite de la réalisation d'une évaluation répétée, il est possible que l'absence d'une période spécifiquement consacrée à cette fin dans le calendrier des activités contribue à faire obstacle à la poursuite d'une analyse approfondie de l'évaluation préalable ou à en limiter la portée.

72. *L'emploi de la même équipe d'évaluateurs pour les deux évaluations.* Il ressort de l'examen de la composition des équipes d'évaluateurs chargées de l'évaluation préalable et de l'évaluation comparative que, dans neuf cas, les équipes avaient été dirigées par la même personne ou bien certains évaluateurs avaient participé aux deux évaluations (dans deux cas le chef de l'équipe responsable de l'évaluation préalable faisait partie de l'équipe chargée de l'évaluation comparative). Or, ces neuf cas comptent parmi les quinze évaluations pour lesquelles le pourcentage des indicateurs ou composantes dont les notes peuvent faire l'objet de comparaisons fiables est élevé. Aucune des sept évaluations comparatives affichant un faible taux de comparabilité des indicateurs ou des composantes n'a été menée par une équipe composée d'au moins certaines personnes

¹⁹ Les NdS/LdM relatives aux douze évaluations préliminaires qui n'ont pas été prises en compte dans cette étude ne stipulent la nécessité de mesurer l'évolution de la performance dans le temps que dans un seul cas (note : trois NdS/LdM ne sont pas disponibles, dont une concerne une auto-évaluation).

ayant procédé à l'évaluation préalable²⁰. Selon trois pays, le recours aux mêmes évaluateurs s'est avéré être « un avantage incontestable » ; « les évaluateurs ont gagné la confiance des autorités et des partenaires de développement » ; et « les autorités avaient autant confiance dans les évaluateurs qu'auparavant » (voir « *Assessing the Impact of the PEFA Framework* » version préliminaire, novembre 2010). Ce qui précède témoigne de l'avantage que présente, en pratique, pour la détermination des progrès dans le temps, le recours à la même équipe d'évaluation pour l'évaluation préliminaire et pour l'évaluation comparative, mais il n'est pas toujours possible de procéder de la sorte. En revanche, si l'emploi de la même équipe d'évaluation peut contribuer à réduire le nombre de changements fallacieux causés par des jugements ou des interprétations (voire même des partis pris) différents, il peut être plus difficile pour une équipe d'admettre que l'évaluation préalable dont elle était responsable était, dans l'ensemble, médiocre²¹. En d'autres termes, il n'est avantageux de recourir à la même équipe d'évaluation que si l'évaluation préalable est jugée être de bonne qualité.

73. Lorsqu'il n'est pas possible d'assurer la participation de quelques-uns des évaluateurs initiaux à l'évaluation répétée, il est néanmoins possible de prendre contact avec le chef d'équipe de l'évaluation préalable pour permettre à l'équipe de l'évaluation répétée d'obtenir des renseignements détaillés (provenant, par exemple, des notes des évaluateurs) sur les notes attribuées dans l'évaluation préalable et de discuter de la validité des changements identifiés au plan de la performance. Le Secrétariat a été partie à quelques contacts de cette nature, qui ont aidé les évaluateurs à comprendre les raisons d'être des notes attribuées aux performances antérieures et à éviter d'émettre des conclusions sur l'évolution des notes sur la base d'hypothèses approximatives. Ceci est particulièrement important lorsque les évaluateurs chargés des évaluations répétées reçoivent des informations qui remettent en question la validité des notes attribuées dans le cadre de l'évaluation préalable.

74. *L'adhésion des autorités à l'évaluation comparative.* L'adhésion des autorités à la préparation de l'évaluation comparative peut être caractérisée de deux manières différentes : i) les autorités « *coopèrent avec l'équipe d'évaluation et la soutiennent* », ce qui implique un rôle relativement passif, et (ii) les autorités « *s'impliquent fortement dans le processus, qu'elles pilotent* », soit un rôle plus actif. Douze pays ont mis sur pied un comité directeur PEFA ou une structure similaire (dont la forme varie selon les pays), pour assurer le suivi, la supervision et l'examen de l'évaluation, ce qui est considéré être une très bonne pratique. L'organisation d'ateliers d'ouverture et de clôture peut aussi contribuer à accroître la qualité de l'évaluation comparative : l'atelier d'ouverture permet de familiariser les autorités avec la méthodologie PEFA et de renforcer leur adhésion à l'exercice PEFA, tandis que l'atelier de clôture permet d'examiner les constatations et de présenter des commentaires. Des ateliers d'ouverture et aussi de clôture ont été organisés dans le cas de 16 évaluations comparatives²².

²⁰ Aucune des douze EP rejetées pour cette étude n'avait eu le même chef d'équipe ou la même équipe d'évaluateurs pour les deux évaluations.

²¹ En fait, l'expérience montre que les évaluateurs sont prêts à admettre que leur évaluation précédente peut être quelque peu déficiente à certains égards et qu'il convient par conséquent de réexaminer les notes attribuées.

²² Les rapports ne fournissent pas tous systématiquement des informations sur l'étendue de la participation du gouvernement et l'organisation d'ateliers ; ils ne sont parfois pas assez précis pour permettre de déduire dans quelle mesure les autorités ont contribué au processus. Il est donc probable que le nombre indiqué soit inférieur à la réalité (il ressort des informations disponibles que 20 ateliers d'ouverture et 19 ateliers de clôture ont été organisés).

75. Sur les 15 évaluations comparatives dont les notes de 80 à 100 % des indicateurs et composantes pouvaient faire l'objet de comparaisons fiables, cinq avaient établi une structure spéciale et cinq avaient organisé à la fois un atelier d'ouverture et un atelier de clôture ; un seul pays a pris ces deux mesures à la fois. En revanche, aucune des sept évaluations comparatives pour lesquelles le taux de comparabilité des indicateurs ou des composantes était inférieur à 50 % n'avait bénéficié de la mise en place d'une structure spécifique pour assurer le suivi du processus²³ et les deux ateliers n'ont été organisés que pour l'une d'entre elles.

76. *Examen d'assurance de la qualité.* Tous les rapports préliminaires des évaluations répétées (y compris les douze qui ne sont pas incluses dans la présente étude) ont été soumis au Secrétariat PEFA pour commentaires, à l'exception de deux auto-évaluations – celles de l'Ouganda-2008 et de la République dominicaine-2009 (qui n'étaient pas des évaluations comparatives). L'examen de la qualité par le Secrétariat PEFA peut être considéré comme un élément important de l'assurance apportée aux parties prenantes concernant la qualité de l'évaluation.

77. **En résumé,** les facteurs intervenant dans le processus qui sont mentionnés ci-dessus ont pu contribuer à la robustesse de la mesure des changements dans le temps, comme indiqué au tableau 4 ci-après. Il semble que lorsque le processus bénéficie de la présence de facteurs positifs, le degré de comparabilité augmente, ce qui n'est pas le cas lorsque ces facteurs sont limités ou absents. Le tableau 4 décompose les degrés de comparabilité des 33 évaluations comparatives en fonction de la présence de facteurs positifs ; les pourcentages sont calculés par rapport au nombre total d'évaluations comparatives, au nombre total d'EC ayant un degré de comparabilité de 80 à 100 % et au nombre total d'EC ayant un degré de comparabilité inférieur à 50 %. Ainsi qu'il a déjà été mentionné, les rapports ne fournissent pas toujours d'informations sur la participation des autorités et sur l'organisation d'ateliers, et celles qui sont présentées sont souvent vagues. Ces deux facteurs sont vraisemblablement plus fréquents que les informations actuellement disponibles ne l'indiquent. À terme, lorsque davantage d'évaluations répétées seront disponibles, il serait utile d'entreprendre des études complémentaires pour avoir une meilleure idée des facteurs qui influent le plus sur la comparabilité.

78. De tous les facteurs concrets considérés, l'examen des versions initiales des rapports par le Secrétariat PEFA est le seul pour lequel il n'a pas été possible d'établir une corrélation avec le degré de comparabilité, pour la simple raison que toutes les EC de l'analyse avaient été examinées par le Secrétariat. Il a cependant été démontré dans de précédents Rapports d'étape que les examens du Secrétariat (ainsi que ceux menés par d'autres parties prenantes) contribuent de façon significative à améliorer le taux de conformité et par conséquent la mesure dans laquelle les notes des indicateurs sont fondées sur des éléments concrets. Les examens du Secrétariat influent donc sur la robustesse de la comparaison des notes des évaluations préalables et des évaluations comparatives.

79. Ainsi que le précise le Secrétariat PEFA dans les « Bonnes pratiques pour l'application du Cadre de mesure de la performance de la GFP » le processus d'évaluation

²³ Les données relatives aux 12 évaluations répétées qui n'ont pas été prises en compte dans la présente étude montrent que, dans deux cas, les autorités avaient mis en place une structure pour gérer l'évaluation PEFA.

« doit être bien maîtrisé pour contribuer à la présentation d'un produit de haute qualité... plus les autorités seront impliquées, plus les agents de l'État bénéficieront de l'exercice ... la compréhension de la méthodologie du Cadre et du processus d'exécution d'une évaluation PEFA par les autorités détermine dans une mesure cruciale leur capacité à diriger efficacement l'opération ». Fait tout aussi important, l'absence d'un processus bien conçu et appliqué peut entraîner la poursuite d'évaluations PEFA suivies de mises à jour et de révisions répétées, qui auront pour effet de compliquer le suivi de l'évolution de la performance dans le temps et d'accroître le risque de faire abstraction des changements intervenus au plan de la performance entre deux évaluations (en Tanzanie, par exemple, l'évaluation PEFA de 2005 a fait l'objet de plusieurs mises à jour et révisions mais l'évaluation de 2010 est la première qui vise de manière explicite à mesurer les progrès).

Tableau 4 : Facteurs concrets et évaluations comparatives

Facteurs concrets	Toutes EC	%	EC comparables à 80-100 %	%	EC comparables à ≤ 50 %	%
Toutes EC	33	100 %	15	100 %	7	100 %
Clarté des NdS/LdM	24	73 %	14	93 %	4*	57 %
Recours à la même équipe d'évaluation	9	27 %	9	60 %	0	0 %
Participation des autorités (structure spécifique)	12	36 %	5	33 %	0	0 %
Ateliers d'ouverture et de clôture	16	48 %	5	33 %	1	14 %
Examen d'assurance de la qualité	33	100 %	15	100 %	7	100 %

- Trois évaluations menées au niveau des administrations infranationales conjointement à une évaluation réalisée au niveau de l'administration fédérale.

80. En conclusion, il ressort de l'examen de 33 évaluations comparatives que la proportion d'indicateurs ou de composantes non notés dans au moins l'une des deux évaluations comparées, qui est plus élevée lorsque les éléments concrets sont insuffisants, a un impact sur le degré de comparabilité. Le degré de comparabilité augmente lorsque les indicateurs ou composantes non notés sont exclus de l'analyse, et ce aussi bien dans l'analyse verticale que dans l'analyse horizontale. Les « problèmes de suivi » qui empêchent de procéder à des comparaisons tiennent principalement aux différences d'interprétation et à la collecte de données sur des éléments nouveaux ayant un impact sur l'évaluation préalable. En outre, certains facteurs concrets peuvent avoir influé sur la qualité de la mesure des changements dans le temps. Il semble que la présence de ces facteurs a contribué à accroître le niveau de comparabilité ; toutefois, ce n'est pas le cas lorsqu'ils n'interviennent que dans une mesure limitée, comme en témoigne la comparaison entre les évaluations affichant un taux de comparabilité élevé et celles dont le taux de comparabilité est faible.

Chapitre 4 – Évolution de la performance de la GFP

4.1 Aperçu

81. Cette section examine les questions suivantes :

- Quels sont les indicateurs qui changent ?
- Est-il possible de détecter des tendances en fonction des catégories d'indicateurs et de caractéristiques nationales ? En fonction des régions ?
- Quels peuvent être les facteurs explicatifs ?

4.2 Constatations

82. L'évolution de la performance a été analysée pour tous les indicateurs ou composantes pour lesquelles il a été possible de procéder à des comparaisons fiables entre deux évaluations successives. Les changements observés représentent donc l'évolution réelle de la performance, et non pas simplement un changement de notes dû à l'obtention de meilleures informations ou à des évaluations ou des interprétations différentes. Si l'on considère l'ensemble des indicateurs ou composantes, on note que 11% ont conservé la note « A », 21 % ont amélioré leur note et 10 % ont conservé la note « D », 9 % ont enregistré une dégradation de leur note, et 21% ont maintenu leur note « B » ou « C » ; enfin, dans 30 %²⁴ des cas, il n'a pas été possible de mesurer les changements de manière valide. Les meilleures et les pires performances ont été enregistrées pour les indicateurs ou composantes suivants :

- PI-24 (ii), *Respect des délais des rapports d'exécution budgétaire produits en cours d'année*, affiche le pourcentage le plus élevé (48 %) de notes « A » dans deux évaluations successives.
- En deuxième position se trouvent PI-3, *Recettes réelles totales par rapport au budget initialement approuvé* (45 %) et PI-15 (ii), *Efficacité du transfert des recettes fiscales au Trésor par l'administration des impôts* (36 %).
- Deux composantes : PI-8 (iii) *Mesure dans laquelle les données budgétaires consolidées (au minimum les données des recettes et des dépenses) sont collectées et communiquées pour l'ensemble des administrations publiques ventilées par secteur*, et PI-28 (i), *Examen en temps opportun des rapports de vérification externe par le pouvoir législatif (pour les rapports reçus au cours des trois dernières années)* ont le pourcentage le plus élevé (39 %) de notes « D » dans deux évaluations successives.
- Vient ensuite PI-23 « *Disponibilité des informations sur les ressources reçues par les unités de prestation de services primaires* (33 %).
- PI-13 (ii), « *Accès des contribuables à l'information sur les obligations fiscales et les procédures administratives* » est la composante enregistrant le pourcentage le plus élevé de notes ayant augmenté d'une évaluation à la suivante (48 %).
- Viennent ensuite trois indicateurs ou composantes, à savoir PI-14 (iii), *Planification et suivi des programmes de contrôle fiscal et d'enquêtes sur la fraude fiscale* (42 %), PI-12 (ii), *Portée et fréquence de l'analyse de viabilité de*

²⁴ Le total dépasse 100 % pour cause d'arrondi.

- la dette et PI-14 (ii), *Effectivité des pénalités pour non respect des obligations d'immatriculation et de déclaration* (38 %).
- Quatre indicateurs ou composantes affichent le pourcentage le plus élevé de notes ayant baissé d'une évaluation à la suivante : PI-23, *Disponibilité des informations sur les ressources reçues par les unités de prestation de services primaires* (21 %), PI-1, *Dépenses réelles totales par rapport au budget initialement approuvé*, PI-3, *Recettes réelles totales par rapport au budget initialement approuvé* et PI-17 (ii), *Degré de consolidation des soldes de trésorerie de l'administration publique* (18 %).

83. Les pays dont les notes sont élevées ou en augmentation (23 pays) sont plus nombreux que ceux dont les notes sont basses ou en diminution (8 pays)²⁵ ce qui témoigne d'une tendance générale à l'amélioration de la GFP dans les pays considérés. Cinq des huit pays affichant les moins bons résultats sont de petits États insulaires dont les difficultés ne sont pas représentatives de celles de l'ensemble de l'échantillon ; ils sont notamment très exposés à des chocs extérieurs comme les ouragans et à la volatilité des recettes du tourisme, se heurtent à de graves problèmes de capacité dus à un exode important de ressources humaines qualifiées, et sont en proie à l'instabilité politique. Les trois autres sont de petits États africains qui sont, pour l'essentiel, confrontés aux mêmes difficultés.

84. Les tendances générales de la performance peuvent être analysées plus en détail grâce à une méthodologie (Andrews, 2009 ; Porter *et al.* 2010) qui catégorise les indicateurs ou composantes sur la base de trois critères principaux (l'analyse « *ne considère que les indicateurs/composantes PI-5 à PI-28 car les indicateurs PI-1 à PI-4 concernent les résultats et la performance des systèmes de GFP et non la qualité même des systèmes de GFP* »)²⁶.

85. Le premier critère établit une distinction entre les composantes PEFA auxquelles la note C ou une note supérieure peut être attribuée par suite de l'adoption d'une nouvelle loi ou de l'annonce d'une nouvelle pratique, même si elle n'est pas mise en œuvre (réforme **de jure** (institutionnelle)) et les composantes pour lesquelles l'obtention de progrès passe nécessairement par la mise en application d'une loi ou le déploiement d'un effort significatif (réforme **de facto** (fonctionnelle)²⁷. Par exemple, dans le cas de la composante à caractère institutionnel PI-11 (i), la note C est attribuée dès lors que le calendrier budgétaire annuel a été établi, même si sa mise en œuvre souffre de retards substantiels et que les entités budgétaires ne disposent pas du temps nécessaire pour produire des estimations détaillées. En revanche, la composante à caractère fonctionnel PI-12 (i) ne peut obtenir la note C que si des prévisions à horizon annuel mobile des agrégats budgétaires couvrant une période de deux ans sont établies.

86. Le deuxième critère permet d'établir une distinction entre les composantes PEFA relatives à la préparation du budget, comme la budgétisation stratégique (prévisions

²⁵ Dans deux pays, le nombre des notes maximales ou plus élevées qu'auparavant était égal au nombre des notes minimales ou faisant état d'une diminution. Les notes maximales et les notes en amélioration ont été regroupées parce que A est la meilleure note qui puisse être attribuée ; les notes minimales ou en diminution sont regroupées parce que D est la note la plus basse qui puisse être attribuée.

²⁶ La liste complète des indicateurs/composantes auxquels un code est affecté se trouve à l'annexe H.

²⁷ Les indicateurs/composantes ont été codés indépendamment par trois spécialistes de la GFP ; Lorsque leurs avis divergeaient, les évaluateurs ont tenu des discussions pour parvenir à un consensus (Andrews, 2009). Voir la liste qui figure à l'annexe H.

pluriannuelles, planification stratégique, planification des investissements, planification de la dette publique) ; la préparation du budget annuel ; l'analyse du budget annuel par le pouvoir législatif ; et la structure des documents budgétaires officiels (actions en **amont**), d'une part, et les composantes relatives à l'exécution du budget telles que la gestion des ressources (notamment la gestion des entrées et sorties de trésorerie, la passation des marchés publics, la gestion de la masse salariale) ; le contrôle interne, l'audit et le suivi internes ; la comptabilité et la communication financière ; l'audit externe ; et l'analyse des rapports d'audit par le Pouvoir législatif (actions en **aval**) d'autre part. Les premières concernent les stades préliminaires du cycle de la planification budgétaire, qui sont observables par les bailleurs et par les investisseurs et qui devraient par conséquent susciter moins de résistance que les secondes, qui portent sur les contrôles et la supervision des dépenses effectives. Par exemple, dans le cas de l'indicateur « en amont » PI-5, la note C peut être attribuée dès lors que la classification budgétaire est conforme aux normes des Statistiques des finances publiques ou à d'autres normes comparables ; or cette conformité résulte d'une pratique institutionnelle qui peut être vérifiée par les bailleurs et les investisseurs. En revanche, les actions « en aval » portent sur des aspects plus délicats de la gestion et du suivi des dépenses effectives. Pour l'indicateur « en aval » PI-7, par exemple, la note C ne peut être attribuée que si les dépenses extrabudgétaires non rapportées ne dépassent pas 5 à 10 % des dépenses totales. Cette condition réduit les opportunités de créer des caisses noires non transparentes dont le montant, dans de nombreux pays, dépasse largement cette limite, et peut donc se heurter à une forte résistance de la part des intérêts en place qui sont les bénéficiaires des dites caisses.

87. Le troisième établit une distinction entre les éléments PEFA qui sont placés sous le contrôle d'organes exécutifs centraux, comme le ministère des Finances (contrôle **concentré**), et ceux qui dépendent de plusieurs ministères ou autorités infranationales (contrôle **déconcentré**). Par exemple, dans le cas de la composante concentrée PI-12 (ii), il suffit, pour qu'une note C soit attribuée, qu'une analyse de viabilité de la dette, au moins pour la dette extérieure, (opération qui peut être accomplie par une petite équipe de techniciens) ait été menée au cours des trois années précédentes. En revanche la composante de contrôle déconcentré PI-12 (iii) exige, pour l'attribution de cette note, que le coût de stratégies sectorielles ait été établi pour plusieurs grands secteurs, ce qui exige la participation de plusieurs entités budgétaires.

88. Bien que ces trois critères donnent lieu à certains chevauchements, ils font ressortir de manière générale les différences entre forme et fonction. En règle générale, les réformes *de jure*, les actions « en amont » et le contrôle « concentré » représentent les caractéristiques institutionnelles du système, tandis que les réformes *de facto*, les actions « en aval » et le contrôle « déconcentré » en représentent les caractéristiques fonctionnelles. L'analyse statique des notes PEFA effectuée par des chercheurs indépendants montre que les éléments institutionnels affichent plus souvent des notes C ou plus hautes que les éléments fonctionnels. L'explication présumée est que, dans le cas des éléments institutionnels, les progrès peuvent résulter de l'adoption d'une loi, d'une réglementation, ou d'une technique nouvelle, n'impliquer qu'un petit nombre de ministères ou se produire aux stades initiaux du cycle budgétaire ; en revanche, pour les éléments fonctionnels, il est plus difficile de réaliser des progrès car il est nécessaire de coordonner les travaux de nombreux ministères, il n'est pas facile de suivre l'exécution

concrète des réformes, celles-ci peuvent remettre en cause des rentes de situation et, partant, susciter une plus grande opposition politique et bureaucratique²⁸.

89. On pourrait, de même, s'attendre à ce que les notes attribuées aux éléments à caractère institutionnel soient très élevées ou en hausse, et que celles des éléments à caractère fonctionnel soient basses ou en diminution. C'est effectivement la tendance générale observée ici. Les éléments à caractère institutionnel sont plus susceptibles d'afficher des notes maximales ou en hausse, tandis que les éléments à caractère fonctionnel sont plus susceptibles d'enregistrer des notes minimales ou en diminution. Il est vrai que l'on observe aussi pour ces derniers un grand nombre de notes très élevées ou en hausse ce qui montre que, même dans les domaines les plus difficiles, il est possible d'afficher d'excellents résultats ou de les améliorer. Il semble, néanmoins, compte tenu du nombre élevé de notes minimales ou affichant la plus forte baisse pour les éléments à caractère fonctionnels, que les progrès dans ces domaines peuvent être fragiles et que des revers sont possibles.

90. Cette méthodologie permet dans une certaine mesure d'expliquer les résultats décrits par les notes présentés ci-dessus. Les trois indicateurs/composantes qui affichent le pourcentage le plus élevé de notes D persistantes présentent les trois caractéristiques fonctionnelles (réformes *de facto*, contrôle déconcentré et actions en aval) qui rendent les progrès relativement difficiles à obtenir. En ce qui concerne les deux indicateurs/composantes concernant la qualité des systèmes de GFP dont la note a diminué dans le pourcentage le plus élevé de cas, l'un (PI-23) présente lui aussi les trois caractéristiques fonctionnelles, tandis que pour l'autre (PI-17 [ii]), les difficultés présentées par la poursuite des travaux en aval du cycle budgétaire surpassent les avantages associés à un contrôle concentré et à des réformes *de jure*. Les deux composantes qui affichent le pourcentage le plus élevé de notes A d'une évaluation à la suivante et les quatre composantes qui ont le pourcentage le plus élevé de notes ayant augmenté, se caractérisent toutes par un contrôle centralisé des organes d'exécution, qui favorise l'obtention de succès rapides, ce qui est un avantage suffisant pour compenser les réformes *de facto* et les actions « en aval » caractéristiques de certains d'entre eux, qui posent plus de difficultés.

91. D'autres aspects de la performance présentent des tendances similaires. Par exemple PI-25(ii), *Respect des délais des états financiers*, est l'indicateur pour lequel le pourcentage de notes passant de D à A (15%) est le plus élevé, et est placé sous le contrôle centralisé d'un organe d'exécution, de même que PI-13 (ii), *Accès des contribuables à l'information sur les obligations fiscales et les procédures administratives*, qui affiche le pourcentage le plus élevé de notes passant de C à A (12%), et PI-14 (iii), *Planification et suivi des programmes de contrôle fiscal et d'enquêtes sur la fraude fiscale*, qui a le pourcentage le plus élevé de notes passant de C à B (24%). Il existe aussi une certaine correspondance entre les bons résultats enregistrés pour ces composantes et les caractéristiques de réformes *de jure* et d'action « en amont », bien qu'elle ne soit pas aussi systématique. En revanche, deux composantes affichent le pourcentage le plus élevé de notes progressant de D à C. La première, PI-26 (i) *Soumission en temps opportun des rapports d'audit au pouvoir législatif*, pour laquelle ce

²⁸ Les qualificatifs « institutionnel » et « fonctionnel » font référence à la caractéristique dominante d'un indicateur ou d'une composante ; certains présentent à la fois des caractéristiques institutionnelles et fonctionnelles. Les qualificatifs attribués n'impliquent aucune recommandation quant à l'ordre dans lequel les réformes doivent être menées, car il peut être nécessaire de lancer des réformes fonctionnelles à un stade précoce.

pourcentage est de 18 %, présente des caractéristiques fonctionnelles (réformes *de facto*, contrôle déconcentré et actions en aval). La deuxième (qui affiche aussi un pourcentage de 18 %), PI-12 (iii), *Existence de stratégies sectorielles avec budgétisation pluriannuelle des dépenses courantes et en capital* a deux caractéristiques fonctionnelles (réformes *de facto* et contrôle déconcentré). Ces exemples montrent que des administrations publiques déterminées peuvent faire progresser des réformes difficiles, peut-être parce que la situation politique est favorable ou qu'ils bénéficient d'un appui plus soutenu des partenaires de développement.

92. Le tableau 5 récapitule la répartition des notes maximales ou en hausse par caractéristique des composantes. Les notes maximales et les notes en amélioration ont été regroupées parce que « A » est la meilleure note qui puisse être attribuée. Les caractéristiques institutionnelles portées dans la colonne de gauche affichent des résultats légèrement meilleurs que les caractéristiques fonctionnelles indiquées dans la colonne de droite ; en effet les éléments assujettis à un contrôle concentré peuvent enregistrer des progrès plus rapides que ceux qui nécessitent la mise en place préalable de lois et de règles formelles pour pouvoir afficher des améliorations fonctionnelles²⁹. Toutefois les différences entre les pourcentages sont plus faibles qu'on n'aurait pu s'y attendre, l'écart le plus réduit étant observé pour les actions en amont et en aval. Cela témoigne clairement de la complexité de chaque cas, la performance étant également fonction des interactions entre le contexte, les interventions des bailleurs, les réseaux de soutien et divers autres facteurs. On observe de nombreuses notes maximales ou en hausse pour les trois domaines fonctionnels de la colonne de droite, ce qui montre que, même dans les domaines les plus difficiles, les notes peuvent être extrêmement hautes ou s'améliorer.

Tableau 5 : Notes maximales ou en hausse, par caractéristique d'indicateur/composante

Caractéristique	% de notes	Caractéristique	% de notes
Contrôle concentré	41	Contrôle déconcentré	26
Réforme <i>de jure</i>	35	Réforme <i>de facto</i>	31
Actions en amont	33	Actions en aval	32

93. Le tableau 6 présente les notes minimales ou en diminution. Les notes minimales ou en diminution sont regroupées parce que D est la note la plus basse qui puisse être attribuée, et la même tendance est visible lorsque l'on n'examine que les notes en diminution. Les données montrent que les caractéristiques fonctionnelles indiquées dans la colonne de droite affichent davantage de notes minimales ou en diminution que les caractéristiques institutionnelles dans la partie gauche du tableau. Cela pourrait signifier que les progrès dans ces domaines peuvent être fragiles et que des revers sont possibles.

94. Il ressort aussi de la comparaison des tableaux 5 et 6 que les caractéristiques institutionnelles enregistrent une proportion beaucoup plus élevée de notes maximales ou en augmentation que de notes minimales ou en diminution, ce qui n'est pas surprenant. Les différences entre les notes des éléments présentant des caractéristiques institutionnelles et fonctionnelles sont plus grandes pour les notes minimales ou en diminution que pour les notes maximales ou en augmentation. Cela signifie que la différence entre les éléments à caractère fonctionnel et les éléments à caractère institutionnel est plus marquée en bas qu'en haut de l'échelle. Il est cependant surprenant

²⁹ La même tendance se dégage lorsqu'on limite l'analyse aux seules notes en amélioration.

que les éléments à caractère fonctionnel affichent une proportion légèrement plus élevée de notes hautes ou en augmentation que de notes minimales ou en diminution, ce qui indique que les réformes produisent des résultats même dans les domaines présentant le plus de difficultés. Cela s'explique peut-être par le fait qu'il peut être plus aisé d'améliorer ce qui fonctionne mal que ce qui fonctionne déjà bien (ce qui est le cas de nombreux éléments à caractère institutionnel). Le test de Student indique que les écarts entre les moyennes sont significatifs au niveau de confiance de 95%³⁰.

Tableau 6 : Notes minimales ou en diminution, par catégorie d'indicateur/composante

Caractéristique	% de notes	Caractéristique	% de notes
Contrôle concentré	9	Contrôle déconcentré	24
Réforme <i>de jure</i>	11	Réforme <i>de facto</i>	23
Actions en amont	15	Actions en aval	19

95. L'examen des notes minimales et de celles qui diminuent le plus suscite produit des résultats préoccupants. Les quatre indicateurs ou composantes dont la note est restée « D » ou a le plus diminué, et qui peuvent être décrits comme des indicateurs de processus, couvrent tous des réformes *de facto* et un contrôle déconcentré, et pour trois d'entre eux également des actions « en aval ». Cela montre clairement à quel point il est difficile non seulement de commencer à faire des progrès, mais aussi de maintenir les avancées réalisées dans ces difficiles domaines fonctionnels de la GFP.

96. L'équipe a également examiné l'évolution des notes entre les sept paires d'évaluations menées à des intervalles plus rapprochés (voir le tableau 7). Alors que l'on s'attendait à ce que les progrès accomplis sur des périodes inférieures à trois à cinq ans ne soient guère perceptibles, il s'avère que le nombre de notes en amélioration est en fait plus élevé pour les évaluations effectuées à des intervalles plus courts, tandis que le nombre de notes non comparables est plus faible. Le pourcentage étonnamment élevé de notes maximales et en augmentation pourrait s'expliquer par le fait que la poursuite d'évaluations comparatives relativement rapidement après l'évaluations préalable a été principalement motivée par le fait que la réalisation d'une EC était une condition nécessaire à l'obtention de l'appui des bailleurs, ce qui a pu inciter les autorités à présenter le système de GFP de la manière la plus favorable possible. Il est aussi possible que les autorités et les bailleurs aient été incités à procéder à une évaluation répétée précoce si les parties prenantes étaient convaincues que l'évaluation répétée indiquerait une amélioration significative des systèmes de GFP.

97. Le Kosovo est un exemple de pays dans lequel la comparaison de deux évaluations rapprochées fait état de rapides progrès : 47 % des indicateurs/composantes ont continué de recevoir la même note ou ont amélioré la note qui leur avait été attribuée dans l'évaluation préalable. Selon le rapport de l'évaluation comparative, ces résultats tiennent à la rapide amélioration du système d'exécution du budget qui est elle-même due à la mise en place d'un système informatique de gestion financière, à l'adoption d'un

³⁰ Il serait utile à un stade ultérieur d'inclure dans les séries les données provenant des 12 évaluations répétées non retenues pour les comparaisons, étant donné que le manque de comparabilité des données en question augmente simplement l'erreur aléatoire associées aux mesures. Il serait ainsi possible de voir si les mêmes tendances ressortent pour de l'échantillon élargi.

plan comptable et à l'établissement d'un compte unique au Trésor. Des améliorations avaient en outre été récemment apportées aux processus d'audit et de contrôle internes ainsi qu'aux audits externes. Le Mozambique est un cas similaire. Selon le rapport de l'évaluation comparative, la plupart des améliorations sont dues aux réformes lancées au niveau de la perception et de la gestion des recettes, de la passation des marchés publics et de la gestion financière par suite du déploiement du système informatique de gestion financière, e-SISTAFE. Certaines des améliorations sont apparues à la suite de l'examen des résultats de l'évaluation préalable, et la conception de nombreuses réformes était déjà bien avancée lors de la réalisation de l'évaluation préalable. Pour autant que ces exemples d'évaluations menées à intervalles rapprochés ne soient pas représentatifs de l'échantillon global, ils montrent que des progrès rapides sont possibles lorsque certains types d'améliorations sont poursuivis dans un environnement favorable.

Tableau 7 : Résultats des évaluations comparatives menées à des intervalles rapprochés et comparaison avec les moyennes globales

	Pourcentage moyen des pays à évaluations rapprochées	Pourcentage moyen global
Maintien de la note « A »	8 %	11 %
Notes en augmentation	31 %	20 %
Maintien de la note « D »	8 %	10 %
Notes en diminution	8 %	8 %
Maintien des notes « B » ou « C »	22 %	21 %
Notes non comparables	24 %	30 %

98. Pour finir, l'équipe a examiné les corrélations pouvant exister entre la mesure de la performance et les caractéristiques nationales. Une étude antérieure (de Renzio 2009) a procédé à une analyse de régression à variables multiples pour comparer les notes PEFA statiques en fonction des caractéristiques nationales. Elle est parvenue à la conclusion qu'il existait une relation positive entre les notes de la GFP et le PIB par habitant, la croissance économique au cours des périodes récentes et un régime démocratique, et une relation négative entre ces notes et la dépendance par rapport aux ressources naturelles et à l'aide, comme on pouvait s'y attendre.

99. En ce qui concerne la mesure de la performance, on pourrait s'attendre également à détecter une relation négative entre le maintien de notes maximales et l'amélioration des notes d'une part, et la dépendance par rapport aux ressources naturelles et par rapport à l'aide, d'autre part. On pourrait aussi s'attendre à ce qu'il existe une relation positive entre ces notes d'une part et le commerce extérieur, le PIB par habitant, la croissance du PIB et la démocratie, d'autre part. Or, l'analyse des corrélations n'a abouti à aucune conclusion significative. Comme indiqué au chapitre 2, il a été décidé de ne pas procéder à une analyse de régression à variables multiples à ce stade en raison de la petite taille de l'échantillon, du domaine de variation limité de la variable dépendante et de la nature subjective des variables explicatives pouvant être retenues. Il serait cependant utile de procéder, à une date ultérieure, à un test dans le cadre duquel les notes de l'évaluation préalable seront incluses comme variable explicative dans l'analyse de régression, de manière à prendre en compte le fait que si les évaluations comparatives des pays développés ne font pas ressortir plus d'améliorations, c'est parce que leurs notes étaient déjà élevées dans l'évaluation préalable. On pourrait aussi analyser les facteurs qui

différencient les 32 pays pour lesquels on dispose d'une évaluation comparative de ceux de l'échantillon beaucoup plus vaste des pays pour lesquels on ne dispose pas d'une telle évaluation. Dans ce cas, on pourrait envisager d'utiliser comme variables explicatives les volumes d'aide (ou, de manière plus restrictive, l'appui budgétaire), la part de l'aide provenant du bailleur le plus important, la part de l'aide fournie par les institutions partenaires de PEFA, la taille du pays, le niveau de démocratie et les indicateurs CPIA, etc.

100. Le fait que les caractéristiques nationales puissent avoir des influences de sens opposés pourrait expliquer l'absence de conclusions significatives : en effet, si l'on peut s'attendre à ce que des pays plus riches, plus capables, démocratiques, en croissance rapide et ouverts aux échanges améliorent leur GFP, il est possible que ces améliorations aient déjà été en grande partie réalisées lors de l'évaluation préalable, ce qui réduit l'envergure possible des améliorations ultérieures. On pourrait de même penser que les pays dépendants des ressources naturelles et de l'aide extérieure soient moins incités à améliorer leur GFP, mais comme les notes attribuées lors de l'évaluation préalable peuvent avoir été très basses, ils peuvent en fait afficher des progrès.

101. En conclusion, les notes attribuées aux éléments à caractère institutionnel de la GFP pour lesquels des progrès peuvent résulter de l'adoption d'une loi, d'une réglementation ou d'une technique nouvelle, ou d'une action ciblée sur un petit nombre d'organismes ou menée à un stade précoce du cycle budgétaire sont plus susceptibles de s'améliorer ou de rester au niveau maximal que celles données à des aspects *fonctionnels* de la GFP pour lesquels l'obtention de progrès passe nécessairement par la mise en application d'une nouvelle loi ou réglementation, ou la coordination des travaux de nombreux organismes, ou encore la poursuite d'une action à un phase plus tardive du cycle budgétaire. La différence est particulièrement prononcée dans les domaines de la GFP où il est possible de réaliser des progrès en ne faisant intervenir qu'une seule ou un petit nombre d'entités publiques, que dans ceux dans lesquels de nombreux organismes sont impliqués. De même, les éléments fonctionnels sont davantage susceptibles d'enregistrer une détérioration de leur note ou de conserver la note minimale que les éléments institutionnels. Quelle que soit la catégorie considérée (institutionnelle ou fonctionnelle), la proportion des éléments affichant des notes maximales ou en hausse est supérieure à celle des éléments affichant des notes minimales ou en diminution, mais les écarts sont plus grands dans la catégorie des éléments institutionnels que dans la catégorie des éléments fonctionnels. L'évolution dynamique des notes PEFA correspond approximativement aux résultats statiques. Toutefois, certains aspects des résultats, comme le fait que les éléments à caractère fonctionnel affichent une proportion légèrement plus élevée de notes hautes ou en augmentation que de notes minimales ou en diminution, sont encourageants et montrent que les réformes produisent des résultats, même dans les domaines présentant le plus de difficultés.

102. Une analyse supplémentaire a été effectuée sur les cas d'intervalles rapprochés, qui montre que des progrès rapides sont possibles lorsque certains types d'améliorations sont poursuivis dans un environnement favorable. Enfin il n'a pas été possible de détecter de relation statistiquement significative avec les différents types de caractéristiques nationales.

103. Les raisons profondes de l'évolution de la performance de la GFP décrite ici — qu'il s'agisse des capacités fournies à titre temporaire par des experts étrangers, de la modification de l'environnement macroéconomique, ou de réformes véritables des

administrations publiques — ne font pas l'objet du présent rapport ; elles pourraient toutefois être examinées dans le contexte de l'étude actuellement consacrée par plusieurs bailleurs à l'évaluation de l'appui aux réformes de la GFP. Les données, de plus en plus nombreuses, provenant des évaluations PEFA répétées seront utiles à cette fin.

Chapitre 5 - Recommandations

104. Sur la base de ces observations, un certain nombre de recommandations sont présentées, dont certaines figurent déjà dans de précédents rapports d'étape. *Le Secrétariat PEFA* devrait prendre les mesures suivantes :

- Revoir les parties des documents existants consacrées au processus d'évaluation (« Liste de contrôle des LdM », « Bonnes pratiques pour l'application du Cadre de mesure de la performance de la GFP », « Note d'information pour les évaluations répétées ») afin de faire ressortir la nécessité : i) d'inclure dans les NdS/LdM une référence spécifique aux notes d'orientation du Secrétariat ; ii) de prévoir des délais et ressources supplémentaires pour l'analyse des changements ; iii) de soumettre les NdS/LdM au Secrétariat pour commentaire ; et iv) de veiller à ce que des informations détaillées sur l'évaluation préalable soient communiquées aux évaluateurs chargés de l'évaluation comparative, en recourant à certains des mêmes évaluateurs (ou, au moins, en sollicitant la collaboration du chef de l'équipe de l'évaluation préalable), et en transmettant aux évaluateurs de l'EC les commentaires des parties prenantes et du Secrétariat PEFA sur l'évaluation préalable.
- Revoir les matériels de formation existants afin de mettre en relief les points précédents.
- Examiner les composantes des indicateurs pour lesquels il est rare de pouvoir procéder à des comparaisons pour déterminer s'il peut être nécessaire d'apporter des précisions au cadre ou de communiquer des directives aux évaluateurs, et s'il convient d'envisager de modifier les normes minimales de notation des composantes.

105. *Les institutions chef de file* devraient veiller au respect d'un certain nombre de bonnes pratiques :

- Les NdS/LdM doivent préciser aussi clairement que possible ce qui est attendu de l'évaluation répétée, comme le recommande la « Note d'information pour les évaluations répétées » du Secrétariat PEFA.
- Les NdS/LdM doivent indiquer de façon suffisamment détaillée aux évaluateurs comment présenter les comparaisons effectuées avec l'évaluation préalable dans le rapport de l'évaluation comparative et prévoir le temps nécessaire pour permettre aux évaluateurs de vérifier la base sur laquelle les notes précédentes ont été attribuées.
- Les NdS/LdM doivent demander à l'équipe d'évaluation d'enregistrer toutes les informations pertinentes (sur un CD par exemple) pour qu'elles soient aisément accessibles et compréhensibles par d'autres experts.
- Les NdS/LdM doivent être inclus dans une annexe de la version préliminaire/finale du rapport d'évaluation.
- Les NdS/LdM doivent être soumis à une procédure d'assurance de la qualité (examen par les pairs) par le Secrétariat PEFA pour garantir le respect en bonne et due forme des points précédents.

- Il importe d'instaurer une procédure pour assurer la communication à l'équipe d'évaluateurs chargée de l'EC de tous les renseignements et documents pertinents se rapportant aux évaluations préalables, notamment le rapport définitif et les commentaires des parties prenantes et du Secrétariat.
- Les institutions chef de file doivent faciliter la prise de contacts entre le chef de l'équipe de l'évaluation préalable et de l'évaluation en cours, même si cela implique des délais supplémentaires et entraîne des coûts additionnels.
- En cas de désaccord avec les commentaires du Secrétariat relatifs aux NdS/LdM et aux évaluations, il importe de communiquer au Secrétariat les raisons de ces désaccords.

106. *Lorsqu'ils procèdent à une évaluation répétée, les évaluateurs doivent :*

- Suivre les directives formulées par le Secrétariat PEFA dans la « Note d'information pour les évaluations répétées ».
- Demander au responsable de l'évaluation (institution chef de file) toutes informations pertinentes relatives à l'évaluation précédente (versions préliminaire et finale du rapport et commentaires formulés dans le cadre de la procédure d'examen de la qualité).
- Demander à l'institution chef de file d'établir des contacts avec l'équipe chargée de l'évaluation préalable.

Annexe A : Évaluations PEFA comparatives

Région	Pays	1 ^{ère} évaluation		1 ^{ère} évaluation répétée		2 ^{ème} évaluation répétée	
		Instit. chef de file	Date du rapport	Instit. chef de file	Date du rapport	Instit. chef de file	Date du rapport
AFR	Burkina Faso	CE	Avr. 07	Autorités	Juin 10		
AFR	Éthiopie	CE	Oct. 07	CE	Oct. 10		
AFR	Éthiopie- Région d'Oromiya	CE	Oct. 07	CE	Juil. 10		
AFR	Éthiopie- Région de Benishangul	CE	Oct. 07	CE	Juil. 10		
AFR	Éthiopie- Région de Harari	CE	Oct. 07	CE	Juil. 10		
AFR	Ghana	Banque mond.	Juin 06	CE	EC		
AFR	Guinée-Bissau	Banque mond.	Juin 06	CE	Mai 09		
AFR	Kenya	DFID	Juil. 06	CE	Mars 09		
AFR	Lesotho	Banque mond.	Juin 07	DFID	Juil. 09		
AFR	Madagascar	CE	Mai 06	Banque mond.	Mai 08		
AFR	Malawi	CE	Juil. 05	CE	Août 06	CE	Juin 08
AFR	Mozambique	CE	Mars 06	Norvège	Fév. 08		
AFR	Ouganda	CE	Mai 06	Banque mond.	Juin 09		
AFR	Sierra Leone	DFID	Déc. 07	DFID	Sept. 10		
AFR	Swaziland	CE	Janv. 07	Banque mond.	Mai 10		
AFR	Tanzanie	Banque mond.	Mai 06	DFID	Mai 10		
AFR	Zambie	DFID	Déc. 05	Autorités	Juin 08		
EAP	Samoa	CE	Oct. 06	Autorités	Avr. 10		
EAP	Timor-Leste	CE	Fév. 07	FMI	Juin 10		
EAP	Tonga	AusAID	Sept. 07	AusAID	Mai 10		
EAP	Vanuatu	CE	Juil. 06	CE	Nov. 09		
ECA	Kosovo	Banque mond.	Mars 07	Autorités	Mai 09		
ECA	Moldova	CE	Avr. 06	Banque mond.	Juil. 08		
ECA	République kirghize	DFID	Janv. 06	SECO	Déc. 09		
ECA	Serbie	Banque mond.	Fév. 07	Autorités	Sept. 10		
LAC	Barbade (la)	CE	Oct. 06	CE	Juil. 10		
LAC	Dominique (la)	CE	Avr. 07	CE	Juin 10		
LAC	République dominicaine	CE	Nov. 07	CE	Sept. 10		
LAC	St-Kitts-et-Nevis	CE	Avr. 07	CE	Déc. 09		
LAC	Ste-Lucie	CE	Oct. 06	CE	Fév. 10		
LAC	Trinité-et-Tobago	CE	Juin 06	CE	Déc. 08		
SAR	Afghanistan	Banque mond.	Déc. 05	Banque mond.	Juin 08		

Annexe B : évaluations PEFA répétées non considérées comme « comparatives »

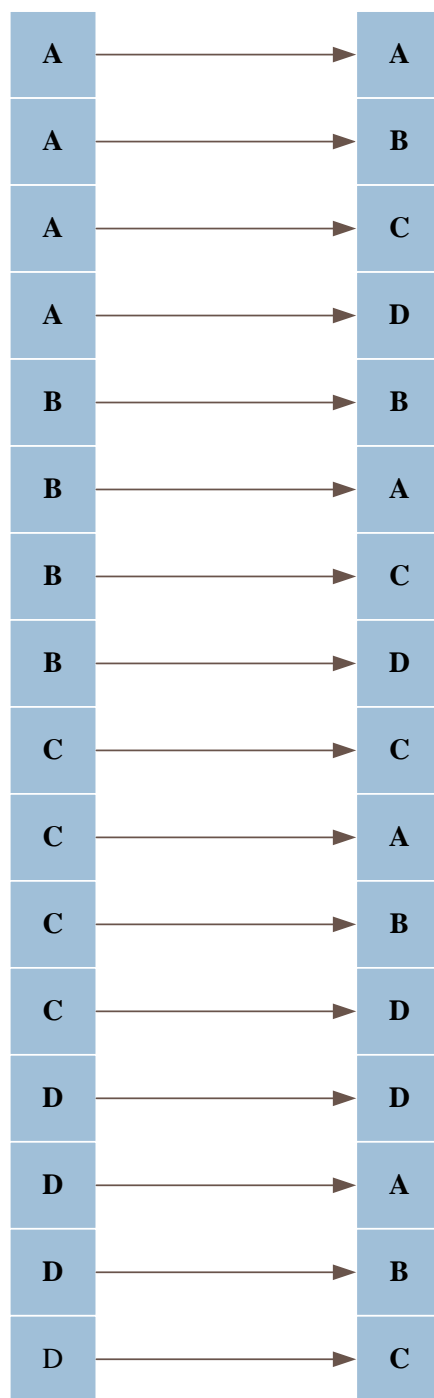
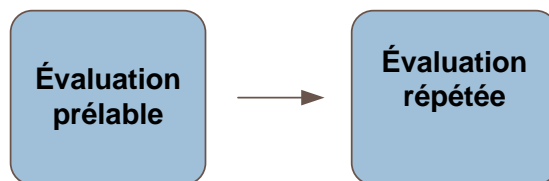
Région	Pays	1 ^{ère} évaluation		Seconde évaluation	
		Institution chef de file	Date du rapport	Institution chef de file	Date du rapport
AFR	Ghana	DFID	Sept. 05	Banque mondiale	Juin 06
AFR	Ouganda	CE	Mai 06	Cour des comptes	Mars 08 (*)
AFR	République centrafricaine	Banque mondiale	Juin 08	CE	Juil. 10
AFR	Sao Tomé-et-Principe	Banque mondiale	Juin 07	CE	Janv. 10
AFR	Togo	Banque mondiale	Jun. 06	CE	Mars 09
EAP	Papouasie-Nouvelle-Guinée	Banque mondiale	Sept. 05	Banque mondiale	Mars 09
EAP	RDP lao	CE	Mai 06	Banque mondiale	Juin 10
LAC	Bolivie	Gouvernement	Oct. 07 (*)	Banque mondiale	Août. 09
LAC	Grenade (la)	CE	Sept. 06	CE	Mars 10
LAC	Honduras	Banque mondiale	Juin 06	CE	Avr. 09
Autres	Montserrat – Royaume-Uni	CE	Sept. 08 (*)	FMI	Déc. 09
LAC	République dominicaine	CE	Nov. 07	Autorités	Nov. 09 (*)

(*) Aucune partie prenante au processus d'évaluation n'a demandé au Secrétariat de procéder à l'examen du rapport préliminaire de l'évaluation PEFA. En conséquence, le Secrétariat PEFA n'a pas communiqué de rapport d'examen ou de commentaires à l'institution chef de file.

- RCA – l'ER ne mesure pas les variations de la performance. La réévaluation des notes présentées dans le tableau n'est pas justifiée.
- Ghana – Le rapport de juin 2006 ne fait aucune référence au rapport de septembre 2005.
- Sao Tomé- et-Príncipe – l'ER ne fait pas référence à l'EP de 2006.
- Togo – l'EP n'était pas conforme au Cadre PEFA (PER/CFAA/PEFA sectoriels). L'ER ne mesure pas les variations de la performance.
- Ouganda – l'évaluation n'a pas fait usage de l'évaluation de 2005 pour mesurer l'évolution de la performance.
- RDP lao – Les notes de l'évaluation partielle PEFA de 2006 n'ont pas été mentionnées ou utilisées dans l'évaluation de 2010.
- Papouasie-Nouvelle-Guinée – le rapport ne tente aucune comparaison avec les notes PEFA précédentes.

- Bolivie – l'ER n'a pas utilisé « l'auto-évaluation » de 2007 pour mesurer les variations de la performance ; elle mentionne l'auto-évaluation de 2007 mais aucune comparaison n'a été envisagée.
- République dominicaine – il s'agit d'une auto-évaluation, présentée comme un exercice de formation pour les fonctionnaires de l'État, antérieure à l'ER de 2010. Elle ne fait aucune référence à l'évaluation antérieure de 2007.
- Grenade – l'ER ne fait aucune référence à l'EP.
- Honduras - PEFA 2009 ne mesure pas l'évolution de la performance sur la durée.
- Montserrat, Royaume-Uni – une évaluation PEFA a été effectuée en 2008 mais il n'y a aucune référence aux notes dans l'évaluation de 2009.

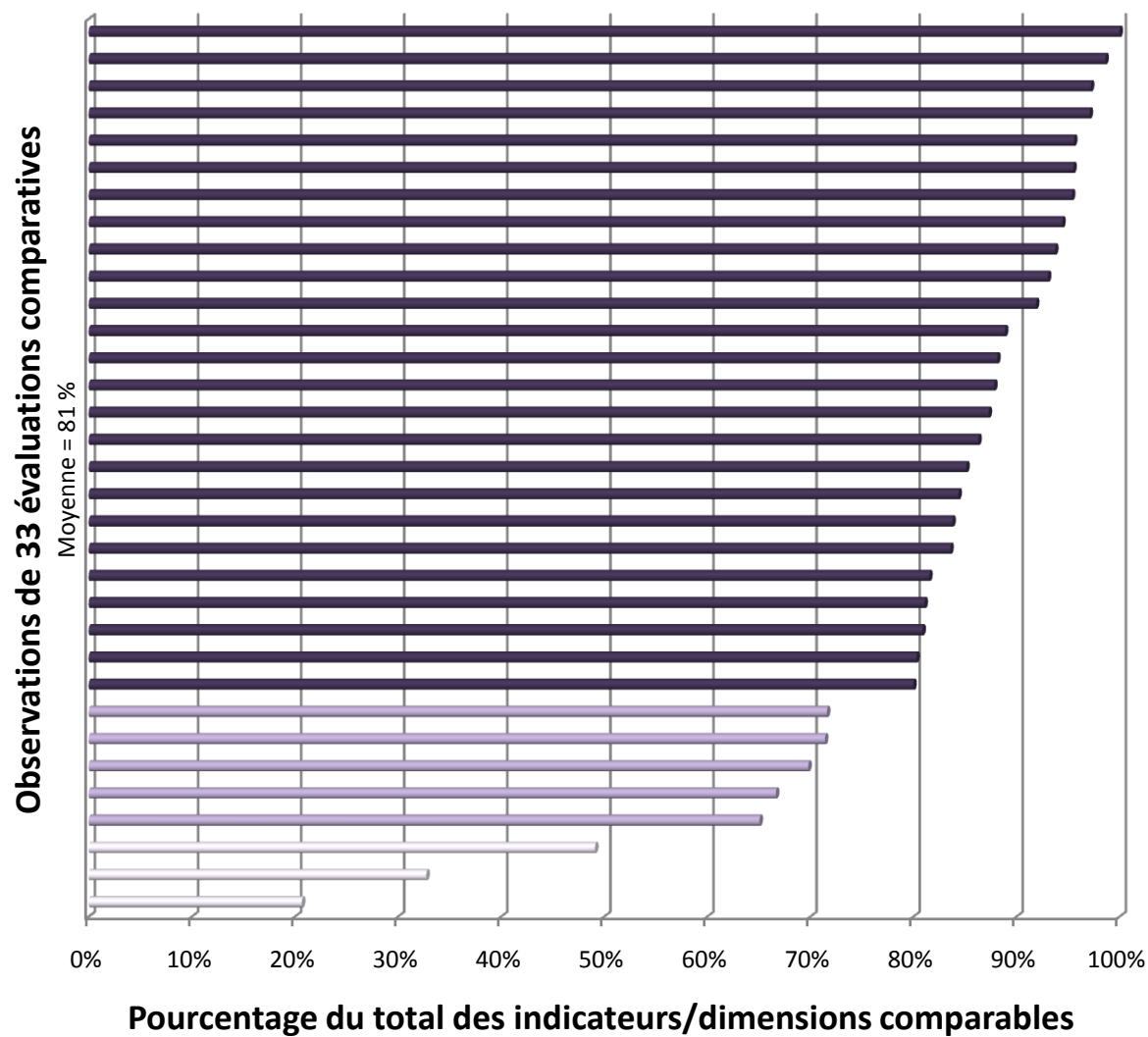
Annexe C : Évolution possible des notes attribuées aux composantes entre l'évaluation préalable et l'évaluation répétée



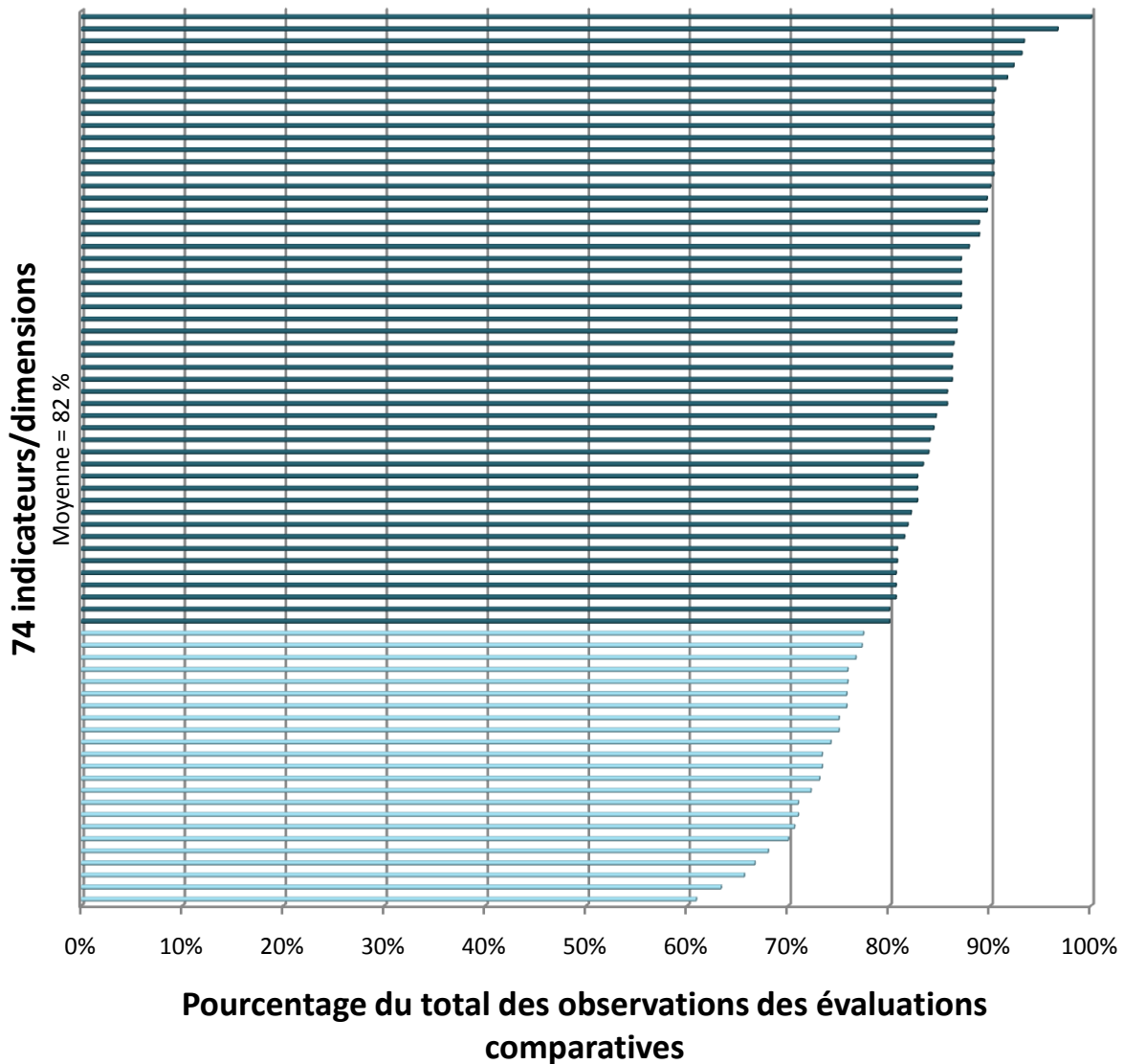
Annexe D : Pays ayant achevé une évaluation de référence avant le 30 juin 2006 (c'est-à-dire il y a plus de 4 1/2 ans) mais qui n'ont pas encore achevé d'évaluation répétée

Pays	Bailleur chef de file	Publié	État d'avancement	Date figurant sur le rapport	État d'avancement actuel de l'évaluation comparative
Bangladesh	Banque mondiale	Oui	Finalisée	Avr-06	En cours
République du Congo	CE	Oui	Finalisée	Mars-06	Pas d'information disponible
Syrie	FMI	Non	Finalisée	Mars-06	Pas d'information disponible
Ouganda-administration locale	CE	Non	Achevée pour l'essentielle	Déc.-05	Prévue
Îles Fidji	BM	Non	Finalisée	Juin-05	Pas d'information disponible

Annexe E : Pourcentage des composantes d'indicateur pouvant faire l'objet de comparaisons fiables entre EC, par évaluation, lorsque les mentions « non notés » sont exclues



Annexe F : Pourcentage des notes pouvant faire l'objet de comparaisons fiables entre EC, pour chaque composante d'indicateur, lorsque les mentions « non notés » sont exclues



Légende³¹:

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

³¹ Les indicateurs sont énumérés par ordre décroissant. Par exemple l'indicateur D-1 (i) est à 100% tandis que PI-20 (iii) est le plus proche de 80%.

Annexe G : Ventilation en pourcentage des indicateurs composantes des EC comparables, non notés et non comparables

Indicateur/ Composante	Comparaison fiable	Non noté	Non comparable
PI-1	64 %	3 %	33 %
PI-2	67 %	6 %	27 %
PI-3	76 %	0 %	24 %
PI-4 (i)	36 %	48 %	15 %
PI-4 (ii)	64 %	21 %	15 %
PI-5	76 %	0 %	24 %
PI-6	82 %	0 %	18 %
PI-7 (i)	42 %	30 %	27 %
PI-7 (ii)	58 %	21 %	21 %
PI-8 (i)	52 %	24 %	24 %
PI-8 (ii)	64 %	24 %	12 %
PI-8 (iii)	67 %	21 %	12 %
PI-9 (i)	64 %	21 %	15 %
PI-9 (ii)	52 %	33 %	15 %
PI-10	88 %	0 %	12 %
PI-11 (i)	85 %	6 %	9 %
PI-11 (ii)	82 %	6 %	12 %
PI-11 (iii)	85 %	6 %	9 %
PI-12 (i)	76 %	6 %	18 %
PI-12 (ii)	73 %	21 %	6 %
PI-12 (iii)	82 %	6 %	12 %
PI-12 (iv)	73 %	6 %	21 %
PI-13 (i)	70 %	6 %	24 %
PI-13 (ii)	85 %	6 %	9 %
PI-13 (iii)	73 %	9 %	18 %
PI-14 (i)	79 %	9 %	12 %
PI-14 (ii)	88 %	9 %	3 %
PI-14 (iii)	82 %	12 %	6 %
PI-15 (i)	39 %	45 %	15 %
PI-15 (ii)	76 %	12 %	12 %
PI-15 (iii)	67 %	18 %	15 %
PI-16 (i)	79 %	6 %	15 %
PI-16 (ii)	85 %	6 %	9 %
PI-16 (iii)	67 %	6 %	27 %
PI-17 (i)	73 %	18 %	9 %
PI-17 (ii)	73 %	12 %	15 %
PI-17 (iii)	73 %	18 %	9 %

Indicateur/ Composante	Comparaison fiable	Non noté	Non comparable
PI-18 (i)	79 %	12 %	9 %
PI-18 (ii)	73 %	12 %	15 %
PI-18 (iii)	73 %	12 %	15 %
PI-18 (iv)	67 %	12 %	21 %
PI-19 (i)	67 %	12 %	21 %
PI-19 (ii)	58 %	9 %	33 %
PI-19 (iii)	67 %	9 %	24 %
PI-20 (i)	67 %	9 %	24 %
PI-20 (ii)	61 %	9 %	30 %
PI-20 (iii)	73 %	9 %	18 %
PI-21 (i)	76 %	12 %	12 %
PI-21 (ii)	79 %	12 %	9 %
PI-21 (iii)	76 %	12 %	12 %
PI-22 (i)	82 %	3 %	15 %
PI-22 (ii)	82 %	6 %	12 %
PI-23	79 %	9 %	12 %
PI-24 (i)	76 %	6 %	18 %
PI-24 (ii)	82 %	6 %	12 %
PI-24 (iii)	82 %	6 %	12 %
PI-25 (i)	76 %	6 %	18 %
PI-25 (ii)	82 %	9 %	9 %
PI-25 (iii)	70 %	9 %	21 %
PI-26 (i)	76 %	9 %	15 %
PI-26 (ii)	64 %	9 %	27 %
PI-26 (iii)	85 %	9 %	6 %
PI-27 (i)	64 %	15 %	21 %
PI-27 (ii)	64 %	15 %	21 %
PI-27 (iii)	73 %	15 %	12 %
PI-27 (v)	70 %	15 %	15 %
PI-28 (i)	85 %	6 %	9 %
PI-28 (ii)	85 %	6 %	9 %
PI-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 %

Annexe H : Ventilation en pourcentage des changements des EC par indicateur et composante

Indicateur/ Composante	A à A	Notes croissantes	D à D	Notes décroissantes	Maintien de la Note B ou C	Non noté/ non comparable
PI-1	18%	18%	6%	18%	3%	36%
PI-2	12%	12%	15%	9%	18%	33%
PI-3	45%	9%	0%	18%	3%	24%
PI-4 (i)	21%	6%	3%	3%	3%	64%
PI-4 (ii)	6%	21%	6%	12%	18%	36%
PI-5	9%	15%	0%	6%	45%	24%
PI-6	15%	36%	0%	0%	30%	18%
PI-7 (i)	24%	3%	3%	9%	3%	58%
PI-7 (ii)	9%	18%	6%	6%	18%	42%
PI-8 (i)	27%	12%	3%	6%	3%	48%
PI-8 (ii)	18%	15%	6%	15%	9%	36%
PI-8 (iii)	12%	6%	39%	6%	3%	33%
PI-9 (i)	0%	12%	3%	9%	39%	36%
PI-9 (ii)	15%	12%	18%	0%	6%	48%
PI-10	3%	27%	3%	9%	45%	12%
PI-11 (i)	21%	9%	0%	15%	39%	15%
PI-11 (ii)	21%	27%	0%	12%	21%	18%
PI-11 (iii)	24%	21%	12%	12%	15%	15%
PI-12 (i)	0%	24%	0%	6%	45%	24%
PI-12 (ii)	18%	39%	3%	6%	6%	27%
PI-12 (iii)	0%	27%	24%	15%	15%	18%
PI-12 (iv)	0%	24%	18%	9%	21%	27%
PI-13 (i)	9%	12%	0%	6%	42%	30%
PI-13 (ii)	15%	48%	0%	3%	18%	15%
PI-13 (iii)	6%	24%	0%	3%	39%	27%
PI-14 (i)	3%	21%	3%	3%	48%	21%
PI-14 (ii)	3%	39%	0%	15%	30%	12%
PI-14 (iii)	0%	42%	3%	6%	30%	18%

Indicateur/ Composante	A à A	Notes croissantes	D à D	Notes décroissantes	Maintien de la Note B ou C	Non noté/ non comparable
PI-18 (i)	21%	18%	15%	15%	9%	
PI-18 (ii)	15%	24%	6%	9%	18%	
PI-18 (iii)	21%	27%	3%	6%	15%	
PI-18 (iv)	0%	21%	6%	12%	27%	
PI-19 (i)	0%	30%	27%	6%	3%	
PI-19 (ii)	0%	15%	3%	12%	27%	
PI-19 (iii)	3%	18%	15%	9%	21%	
PI-20 (i)	3%	24%	6%	9%	24%	
PI-20 (ii)	0%	18%	6%	3%	33%	
PI-20 (iii)	0%	21%	3%	6%	42%	
PI-21 (i)	6%	21%	27%	0%	21%	
PI-21 (ii)	0%	33%	12%	12%	21%	
PI-21 (iii)	3%	18%	27%	12%	15%	
PI-22 (i)	6%	33%	6%	6%	30%	
PI-22 (ii)	6%	30%	15%	6%	24%	
PI-23	6%	9%	33%	21%	9%	
PI-24 (i)	21%	18%	0%	12%	24%	
PI-24 (ii)	48%	21%	0%	3%	9%	
PI-24 (iii)	21%	24%	3%	9%	24%	
PI-25 (i)	12%	18%	9%	15%	21%	
PI-25 (ii)	30%	27%	9%	6%	9%	
PI-25 (iii)	12%	21%	6%	3%	27%	
PI-26 (i)	0%	30%	12%	9%	24%	
PI-26 (ii)	0%	30%	9%	12%	12%	
PI-26 (iii)	3%	24%	15%	9%	33%	
PI-27 (i)	9%	15%	0%	6%	33%	
PI-27 (ii)	6%	24%	6%	6%	21%	
PI-27 (iii)	18%	15%	15%	9%	15%	

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

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

Annexe I : Méthodologie de codage de la GFP
(établie par M. Andrews)

Composante PEFA	De jure	De facto	Concentré	Déconcentré	Amont	Aval
PI-5	1	0	1	0	1	0
PI-6	1	0	1	0	1	0
PI-7i	0	1	0	1	0	1
PI-7ii	0	1	0	1	0	1
PI-8i	1	0	1	0	1	0
PI-8ii	0	1	0	1	1	0
PI-8iii	0	1	0	1	0	1
PI-9i	0	1	0	1	0	1
PI-9ii	0	1	0	1	0	1
PI-10	1	0	1	0	1	0
PI-11i	1	0	1	0	1	0
pi-11ii	1	0	1	0	1	0
PI-11iii	0	1	0	1	1	0
PI-12i	0	1	1	0	1	0
PI-12ii	1	0	1	0	1	0
PI-12iii	0	1	0	1	1	0
PI-12iv	0	1	0	1	1	0
PI-13i	1	0	1	0	0	1
PI-13ii	1	0	1	0	0	1
pi-13iii	1	0	1	0	0	1
PI-14i	1	0	1	0	0	1
PI-14ii	0	1	1	0	0	1
PI-14iii	1	0	1	0	0	1
PI-15i	0	1	0	1	0	1
PI-15ii	0	1	1	0	0	1
PI-15iii	0	1	1	0	0	1
PI-16i	1	0	1	0	0	1
PI-16ii	0	1	0	1	0	1
PI-16iii	0	1	0	1	0	1
PI-17i	1	0	1	0	0	1
PI-17ii	1	0	1	0	0	1
PI-17iii	1	0	1	0	0	1
PI-18i	0	1	0	1	0	1
PI-18ii	0	1	0	1	0	1
PI-18iii	1	0	0	1	0	1
PI-18iv	0	1	0	1	0	1

PI-19i	0	1	0	1	0	1
PI-19ii	0	1	0	1	0	1
PI-19iii	1	0	0	1	0	1
PI-20i	0	1	0	1	0	1
PI-20ii	1	0	0	1	0	1
PI-20iii	0	1	0	1	0	1
PI-21i	0	1	0	1	0	1
PI-21ii	1	0	0	1	0	1
PI-21iii	0	1	0	1	0	1
PI-22i	0	1	1	0	0	1
PI-22ii	0	1	0	1	0	1
PI-23	0	1	0	1	0	1
PI-24i	1	0	1	0	0	1
PI-24ii	0	1	1	0	0	1
PI-24iii	0	1	1	0	0	1
PI-25i	0	1	0	1	0	1
PI-25ii	0	1	1	0	0	1
PI-25iii	1	0	1	0	0	1
PI-26i	0	1	0	1	0	1
PI-26ii	0	1	0	1	0	1
PI-26iii	0	1	0	1	0	1
PI-27i	1	0	0	1	1	0
PI-27ii	1	0	0	1	1	0
PI-27iii	1	0	0	1	1	0
PI-27iv	1	0	0	1	1	0
PI-28i	0	1	0	1	0	1
PI-28ii	0	1	0	1	0	1
PI-28iii	0	1	0	1	0	1
Total	26	38	26	38	16	48

Annexe J : Bibliographie

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

de Renzio, Paulo, 2009. Taking Stock: What do PEFA Assessments tell us about GFP systems across countries? Working Paper 302. London: Overseas Development Institute, accessed May 11, 2011 <http://siteresources.worldbank.org/PEFA/Resources/TakingStockRenzio2007.pdf>

Department for International Development, 2009. Fiduciary Risk Assessment. London: DFID.

Government of the Republic of Bangladesh and Department for International Development, 2007. Assessment of the Impact of Financial Management Reforms in Bangladesh 1992 to 2006.

Mackie, Andrew and Caprio, Giovanni, 2010. Assessing the Impact of the Cadre PEFA, A study for the PEFA SC, draft report, November.

Porter, Doug, Matt Andrews, Joel Turkewitz and Clay Wescott, 2010. Managing Public Finance and Procurement in Fragile and Conflicted Settings. Washington: World Bank, accessed May 11, 2011, <http://wdr2011.worldbank.org/procurement>

World Bank, 2006. Bangladesh Country Assistance Strategy 2006-2009. Washington: World Bank, accessed May 11, 2011, http://siteresources.worldbank.org/BANGLADESHEXTN/Resources/CAS_MAIN_BOOK_FINAL.pdf

World Bank, 2007. Public Sector Accounting and Auditing Gap Analysis. Washington: World Bank

World Bank, 2009. Bangladesh Public Expenditure and Institutional Review. Washington: World Bank, accessed May 11, 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)