



Bonnes pratiques dans la conduite d'une évaluation répétée

Note d'orientation à l'intention des commanditaires et des évaluateurs

Version finale

Secrétariat du PEFA
1^{er} février 2010

Bonnes pratiques dans la conduite d'une évaluation répétée

Résumé analytique de l'évaluation

Depuis sa mise en place en 2005, le Cadre de mesure de la performance PEFA (Dépenses publiques et responsabilité financière) a été utilisé dans plus de 150 évaluations du système de gestion des finances publiques (système de GFP) d'un pays. Bien que bon nombre des évaluations récentes soient des « évaluations répétées », souvent elles ne répondent pas à l'un des principaux objectifs du Cadre, à savoir donner une vision claire des évolutions spécifiques de la performance depuis la première évaluation.

C'est pourquoi il a été demandé au Secrétariat de proposer des orientations sur les « bonnes pratiques ». La présente note formule des suggestions pour chacune des étapes d'une évaluation répétée type.

En résumé, les orientations sont structurées comme suit :

À L'ATTENTION DES COMMANDITAIRES / PROMOTEURS DE L'ÉVALUATION

NOTE DE FAISABILITE / TERMES DE RÉFÉRENCE

Ne pas faire d'évaluations trop fréquemment : les changements mettent du temps à produire des effets ; l'intervalle recommandé avant une utilisation répétée du cadre d'analyse du PEFA est **au minimum de trois ans**.

- S'assurer que les **notes de faisabilité / termes de référence sont suffisamment détaillés** et que toutes les parties prenantes les comprennent bien.
- **Faciliter l'accès à la documentation** relative à l'évaluation précédente, publiée ou non, aux notes non publiées et, s'il y a lieu, aux contributions orales du responsable de l'équipe d'évaluation précédente.
- **L'organisme chef de file de l'évaluation précédente** (s'il diffère de celui qui mène la nouvelle évaluation) doit faire partie du groupe de référence de la nouvelle évaluation.
- Il serait **avantageux d'utiliser la même équipe** que celle de l'évaluation précédente, pourvu que cette équipe ait fait un rapport de qualité et que les membres de l'équipe n'aient pas participé dans l'intervalle à la mise en place et à l'exécution de la réforme de la GFP du pays, cette participation pouvant constituer un conflit d'intérêt lors de la nouvelle évaluation.

À L'ATTENTION DES ÉVALUATEURS

PRÉPARATION

- **Contactez le responsable de l'équipe d'évaluation précédente** pour obtenir les informations complémentaires voulues sur les données et les sources (y compris les observations formulées par les évaluateurs-experts).
- **Ne pas aborder l'évaluation en pensant *a priori* que les évaluateurs précédents se sont trompés dans leur notation ou au contraire qu'ils avaient entièrement raison !**
- **Prendre l'évaluation précédente comme point de départ** pour expliquer l'évolution de la performance du système de GFP (plutôt que d'examiner chaque indicateur indépendamment de l'évaluation précédente puis de comparer les notations, approche risquant de ne pas repérer les différences de définitions, etc.).

Bonnes pratiques dans la conduite d'une évaluation répétée

TRAVAIL SUR LE TERRAIN

- **Vérifier les éléments sur lesquels repose l'attribution de la notation précédente** et repérer tous les indicateurs pour lesquels on ne dispose pas d'informations suffisantes pour cette vérification.
- **Conserver des notes complètes sur les données et les sources** pour chaque notation en vue du suivi à venir (même si ces données et ces sources doivent figurer dans un dossier confidentiel qui ne sera pas repris dans le rapport proprement dit).
- Ne pas utiliser la notation « D » en cas d'informations insuffisantes pour attribuer une notation.
- **Ne pas chercher à modifier les notations attribuées lors de l'évaluation antérieure** : ces notations figurent dans des documents – publiés ou non – que diverses parties prenantes auront utilisés.

RÉDACTION DU RAPPORT PROVISOIRE

- Au cas où des erreurs seraient constatées dans les notations antérieures, les expliquer dans l'analyse des changements apportés aux scores attribués aux indicateurs. **On peut insérer une note dans le tableau de présentation générale des indicateurs pour expliquer que les notations actuelle et précédente ne sont pas comparables.**
- **Etre précis quant aux sources des données** (cela aidera les prochains évaluateurs).
- **Inclure la note de faisabilité / les termes de référence en annexe** au rapport.
- Dans le résumé analytique de l'évaluation :
 - **Donner un aperçu général des changements apportés aux notations de la performance** par rapport au précédent rapport et établir un tableau avec les scores n, puis n+1;
 - **Enumérer les indicateurs** qui se sont dégradés, ceux qui se sont améliorés et ceux qui n'ont pas changé ;
 - Ne pas regrouper les résultats de n à n+1 pour obtenir une amélioration en % !
- Dans la description narrative :
 - Donner un **aperçu général des problèmes que posent les scores précédents** en fonction des données ou des modifications de définitions ;
 - **Déterminer l'incidence des désaccords** par rapport aux scores de l'évaluation précédente ;
 - **Exposer les raisons des changements apportés aux scores entre les deux évaluations, indicateur par indicateur**, afin que le lecteur saisisse en détail ce qui s'est passé.

1 Objet de la présente note

Depuis 2005, le Cadre de mesure de la performance PEFA a été utilisé pour évaluer les systèmes de GFP dans plus d'une centaine de pays. Jusqu'à présent, la plupart des évaluations ont été des évaluations initiales mais comme on pouvait s'y attendre, les évaluations répétées sont de plus en plus fréquentes, puisque près de trois années se sont déjà écoulées dans certaines régions depuis la conduite des évaluations initiales.

Bien qu'une partie croissante des évaluations soit des évaluations répétées, souvent toutefois elles ne répondent pas à l'un des principaux objectifs du Cadre, à savoir donner une vision claire des évolutions spécifiques de la performance du système de GFP depuis l'évaluation précédente. Dans

Bonnes pratiques dans la conduite d'une évaluation répétée

certain cas, les progrès sont décrits dans le Résumé analytique de l'Évaluation sans pour autant être détaillés pour chacun des indicateurs.

Le nombre toujours croissant d'évaluations répétées incite les évaluateurs, les responsables du secteur public et les représentants des agences de développement à solliciter au Secrétariat des orientations en matière de « bonnes pratiques » pour les évaluations répétées, d'où la rédaction de la présente note.

2 Contexte : utilisation du Cadre

Le Cadre de mesure de la performance de la GFP est l'un des trois éléments de l'*Approche renforcée* pour soutenir les réformes de la GFP : ces réformes sont les suivantes : (i) élaboration d'une stratégie de réforme et d'un plan d'action sur la GFP sous la direction du pays ; (ii) mise en place d'un programme d'action pour la GFP pluriannuel coordonné par les partenaires au développement, qui soutient la stratégie nationale de réforme des systèmes de GFP et s'y conforme ; (iii) et constitution d'une base d'information commune.

Le Cadre est un outil conçu pour atteindre le troisième objectif – une base d'information commune –, en particulier pour mesurer la performance de la GFP de pays à divers stades de développement au fil du temps, en autorisant la constitution d'une base de référence permettant d'évaluer les progrès.

Les parties prenantes peuvent s'attendre à ce que l'utilisation répétée du Cadre, si elle est faite de manière cohérente, atteste de la mesure dans laquelle le pays améliore ou non sa performance en matière de GFP. En outre, le Rapport de performance de la GFP reconnaîtra les efforts déployés par un pays pour réformer son système de GFP en décrivant les réformes récentes et en cours, même si celles-ci n'ont peut-être pas encore eu d'effet mesurable sur la performance de la GFP.

3 Cohérence

À l'évidence, deux évaluations menées à un certain intervalle de temps ne seront comparables que si elles appliquent la méthodologie de manière cohérente. A supposer que l'évaluation précédente offre une application rigoureuse de la méthodologie, les parties prenantes voudront dans l'idéal s'assurer de la rigueur du suivi des progrès dans le temps, en recourant :

- AUX MÊMES évaluateurs ;
- AU MÊME promoteur ;
- AUX MÊMES contreparties ;
- AUX MÊMES sources d'information.

Il est toutefois peu probable que l'ensemble de ces éléments, ou ne serait-ce que l'un d'entre eux, puisse être utilisé pour une évaluation répétée. Si, par exemple, des consultants étaient employés pour l'évaluation précédente, la même équipe est peu susceptible d'être disponible trois ans plus tard, et ce en raison de la nature même du travail de consultants. Pareillement, certains responsables du secteur public auront certainement changé de poste après plusieurs années. En outre, moins de la moitié des évaluations répétées ont été parrainées par la même agence de développement. Certaines sources

Bonnes pratiques dans la conduite d'une évaluation répétée

d'information, comme les observations des commanditaires et des pairs (dont le Secrétariat) sur les rapports provisoires de l'évaluation précédente, peuvent enfreindre les obligations de confidentialité.

L'évaluation antérieure peut aussi ne pas avoir été considérée comme étant de bonne qualité, ce qui soulève une autre question : cette évaluation constitue-t-elle la base fiable tellement indispensable à un suivi pertinent des progrès et, dans la négative, comment doivent procéder les nouveaux évaluateurs?

Ces problèmes pratiques soulignent l'importance cruciale de la question de savoir ce qu'est une « bonne pratique » pour les évaluateurs qui entreprennent une évaluation répétée.

4 Problèmes rencontrés dans les évaluations répétées

Une évaluation répétée a pour principal objectif de suivre l'évolution de la performance depuis l'évaluation précédente. Pour ce faire, il faut s'assurer que l'on compare des choses comparables. En d'autres termes, il faut comprendre les éléments de justification sur lesquels reposent les notations de l'évaluation précédente. Cela n'est pas toujours évident. À titre d'exemple, les chiffres des recettes et dépenses dont disposaient les évaluateurs de l'évaluation précédente peuvent avoir changé après avoir été vérifiés et les évaluateurs de l'évaluation répétée peuvent avoir obtenu des données budgétaires différentes pour le même exercice. Ce point et d'autres problèmes sont explicités ci-après.

L'examen des plus de vingt évaluations répétées effectuées à ce jour appelle immédiatement deux remarques. En premier lieu, toutes ces évaluations, ou presque, ont été réalisées après moins de temps que le délai suggéré par le Cadre de Mesure de la Performance PEFA (soit 3-5 ans) alors même qu'il est généralement admis que les tentatives de réforme mettent un certain temps à produire leurs effets. En second lieu, ces nouvelles évaluations sont de qualité variable : certaines de bonne qualité, d'autres non. Plusieurs raisons peuvent expliquer ces faiblesses :

- Premièrement, bien que consciente de procéder à une évaluation répétée, l'équipe d'évaluation peut avoir pris la décision délibérée de ne pas consulter le rapport précédent, comme en témoigne cette citation extraite d'un rapport d'évaluation : « *L'équipe voulait tirer des conclusions en toute indépendance, sans être influencée par les évaluations précédentes. Ce point a été discuté avec les parties prenantes au cours de l'évaluation* ». Dans cette situation, les modifications apportées aux notations deviennent bien plus difficiles à expliquer.
- Deuxièmement, la note de faisabilité / les termes de référence peuvent ne pas avoir suffisamment précisé la mission des évaluateurs en matière de suivi des progrès.
- Troisièmement, les évaluateurs n'ont peut être pas pris connaissance de l'évaluation précédente, et si celle-ci a été parrainée par un organisme chef de file différent, n'ont peut être pas eu accès aux projets de document ou même au rapport final si ce dernier n'a pas été rendu public ; les évaluateurs peuvent donc estimer qu'ils ne sont pas en mesure de faire une comparaison.
- Quatrièmement, les équipes d'évaluation peuvent ne pas avoir accès aux observations des pairs (y compris le Secrétariat) du *projet* de rapport de l'évaluation précédente, qui a peut être pointé des problèmes de notation alors **non** pris en compte dans la version *finale* du rapport. Ainsi, les évaluateurs ne disposent peut-être pas de toutes les informations nécessaires pour tenter d'établir une comparaison.

Bonnes pratiques dans la conduite d'une évaluation répétée

- Cinquièmement, les équipes d'évaluation peuvent ne pas s'accorder sur le bien-fondé du score attribué par l'évaluation précédente, et ainsi ne pas être en mesure de faire une comparaison d'une réelle utilité.
- En dernier lieu, certaines évaluations répétées ont pour objectif d'établir une base de référence plus exacte que celle de l'évaluation précédente, auquel cas la comparaison peut paraître dénuée d'intérêt.

5 Exemples de bonnes pratiques

Pour le Secrétariat, une évaluation répétée est satisfaisante lorsqu'elle confirme la véracité des éléments sur lesquels repose l'attribution de la notation antérieure et lorsqu'elle identifie les indicateurs pour lesquels on ne disposait pas d'informations suffisantes pour cette confirmation. En outre, une évaluation répétée satisfaisante doit prendre en compte toutes les erreurs manifestes commises dans l'utilisation des éléments de justification sur lesquels repose la notation précédente afin de comparer ce qui est comparable.

Voici des exemples de faiblesses courantes relevées par les observations du Secrétariat lors de l'examen de projets de rapports récents :

- *« Une attention insuffisante est portée au suivi des progrès. Le rapport n'expose pas les facteurs explicatifs de progrès au titre de chaque indicateur, s'ajoutant aux notations comparatives. »*
- *« L'absence de comparaison entre les notations de l'année 200X et de l'année 200Y par indicateur (voire par sous-indicateur) empêche de comprendre les changements (ou l'absence de changements). »*
- *« La comparaison ou l'explication n'est fournie que pour quelques indicateurs. »*
- *« Il n'y a pas de tableau récapitulatif indiquant les notations antérieure et actuelle et expliquant les progrès ou l'absence de progrès. »*

Les types de problèmes susmentionnés seront résolus si :

- Le résumé analytique de l'évaluation présente un bref aperçu de l'évolution des notations depuis le rapport précédent, ainsi qu'un tableau indiquant les deux séries de notations ;
- les raisons de l'évolution des scores entre les deux évaluations sont présentées, pour chaque indicateur ;
- une présentation détaillée des progrès accomplis entre les deux évaluations figure en annexe.

Les scores attribués aux indicateurs donneront une idée sommaire des évolutions dans le temps, mais les aspects individuels peuvent évoluer différemment et la performance peut ne pas changer suffisamment pour modifier le score attribué à tel ou tel indicateur (auquel cas il peut être approprié d'utiliser une flèche). Une explication plus détaillée sera alors nécessaire.

Pour illustrer la bonne pratique, le Secrétariat propose les exemples suivants tirés de rapports récents :

Bonnes pratiques dans la conduite d'une évaluation répétée

- dans le Résumé analytique de l'évaluation :

.... « Il ressort du tableau A1 que 7 indicateurs ont enregistré une dégradation depuis 2005. De fait, certaines des notations de 2005 paraissent actuellement surévaluées, ainsi que l'expliquent les tableaux comparatifs tout au long du rapport. Un seul indicateur, le D-1, paraît s'être réellement dégradé. Cette dégradation est surprenante au vu des modèles du Programme de réformes et doit faire l'objet d'un suivi dans le cadre de l'examen. Douze indicateurs apparaissent inchangés, bien que pour certains (PI-10, 15, 18, 26 et 27) il y ait eu d'importantes réformes, quoiqu'insuffisantes pour modifier leurs notations. Douze autres indicateurs apparaissent en progrès, mais une réévaluation de l'indicateur PI-5 en 200X ramène leur nombre à 11 (PI-4, 6, 9, 13, 14, 17, 20, 21, 22, 24, et D-2). Pour certains, cela résulte de la mise en place du nouveau système de GFP et de son utilisation toujours plus fréquente, ainsi que d'autres réformes. Il reste toutefois à remédier plus efficacement à plusieurs points faibles, comme le montre le résumé. »

Tableau A1

		Notation 200X	Notation 200Y
PI-1	Dépenses réelles totales par rapport au budget initialement approuvé	B	B
PI-2	Composition des dépenses réelles par rapport au budget initialement approuvé	C	C
PI-3	Recettes réelles totales par rapport au budget initialement approuvé	A	A
PI-4	Stock et suivi des arriérés de paiement de dépenses	D	D+

- pour les indicateurs individuels :

« Comparaison de 2004 et 2006 – il ressort des informations disponibles en 2007 que la notation attribuée à l'indicateur PI-1 en 2004 était de fait erronée, en raison particulièrement de la comptabilisation des dépenses relatives au service de la dette dans le calcul. Les écarts alors enregistrés par rapport aux dépenses prévues étaient de +3,7 % en 2002, 10,8 % en 2003 (en raison des inondations qui ont ravagé le pays en 2000 et 2001, nécessitant de grosses dépenses au titre des secours et de la reconstruction qu'il était difficile de budgétiser avec précision) et 3,7 % en 2004. Notre analyse mise à jour des données pour l'année 2004 (à partir des états financiers définitifs) indique un écart de 5,19 % pour l'exercice 2004, ce qui aurait donné un score « B ».

- Ensuite, une annexe offrant un aperçu détaillé de la notation PEFA n et PEFA n+1 pour chaque indicateur, ainsi qu'une description des progrès accomplis entre n et n+1 :

		Notation 200X	Notation 200Y	Evolution entre 0X et 0Y
PI-1	Dépenses réelles totales par rapport au budget initialement approuvé	C	B	L'attention continument portée à la discipline budgétaire a permis d'améliorer la prévisibilité de l'ensemble du budget

Bonnes pratiques dans la conduite d'une évaluation répétée

PI-2	Composition des dépenses réelles par rapport au budget initialement approuvé	D	D	L'écart moyen entre dépenses totales budgétées et dépenses totales réelles des MDA s'est légèrement amélioré, ce qui peut s'expliquer à la fois par un financement soutenu des programmes prioritaires et une meilleure mise à disposition des ressources allouées aux MDA tout au long de l'année. L'écart par rapport au budget n'en demeure pas moins important.
PI-3	Recettes réelles totales par rapport au budget initialement approuvé	A	A	La prudence continue des prévisions de recettes et les mesures coercitives renforcées ont maintenu des niveaux élevés de collecte de recettes par rapport aux montants budgétés.
PI-4	Stock et suivi des arriérés de paiement de dépenses	D+	B+	
(i)	<i>Stock d'arriérés de paiement de dépenses et évolution récente du stock</i>	D	B	La stratégie quinquennale des pouvoirs publics de réduction des arriérés de paiement de dépenses a permis une diminution du total des arriérés d'un exercice sur l'autre.
(ii)	<i>Disponibilité des données pour le suivi du stock d'arriérés de paiement de dépenses</i>	B	A	Des rapports trimestriels sont maintenant régulièrement établis par le contrôleur financier de l'audit interne.

6 Désaccord avec les notations précédentes

Pour pouvoir réellement suivre et saisir les évolutions de la performance de la GFP et se fier aux comparaisons des scores attribués à un indicateur dans le temps, les lecteurs doivent comprendre ce qui a changé et l'ampleur de ce changement via la description narrative du rapport de performance de la GFP.

Il faut aussi garder à l'esprit que tous les changements de notation ne représentent pas nécessairement un changement dans la performance du système de GFP. Les raisons possibles de ces changements (extraites d'évaluations récentes) sont notamment les suivantes :

- **Accès différent à l'information**

Exemple : PI-7 : « Cette fois, on a pris en compte une étude des opérations quasi-budgétaires d'entreprises paraétatiques. Or cette étude n'était pas disponible pour l'évaluation antérieure.

Bonnes pratiques dans la conduite d'une évaluation répétée

Aussi la notation est-elle plus faible cette fois-ci sans qu'il y ait de preuve attestant une quelconque dégradation de la performance ».

Autre exemple : PI-4: « Cette fois-ci, l'évaluation se fonde sur une définition différente des arriérés. La dernière fois, l'évaluation se basait sur la définition comptable et comprenait donc seulement les paiements qui avaient été acceptés et traités pour être effectués par le Trésor. Les arriérés en volume ont donc augmenté sans pour autant que la performance se soit nécessairement dégradée. »

- **Différences d'échantillonnage**

Exemple PI-21 : « Dans un rapport précédent, l'évaluation était fondée sur la conjoncture telle que décrite par les Ministères de l'agriculture, de la justice et des finances. Cette fois-ci, l'évaluation est basée sur des informations provenant des Ministères de l'éducation et de la justice. Il n'est donc pas certain qu'une comparaison directe soit valable ».

- **Interprétation différente**

Exemple PI-24(ii) : « Lors de l'exercice précédent, les rapports sur l'exécution du budget ont été publiés 4 à 7 semaines après la fin du trimestre, 5,5 semaines en moyenne : on a attribué une notation « B ». Pour la dernière évaluation, la performance était exactement la même mais avec une notation « C » car deux rapports avaient un retard de plus de 6 semaines. »

- **Méthode de notation – des erreurs peuvent se produire**

Il faut expliquer l'incidence des erreurs sur les changements de notations. Quelques autres exemples :

- attribution erronée d'une notation en dépit d'éléments d'information très probants ;
- utilisation de la méthode M1 d'agrégation des composantes là où l'on aurait dû recourir à la méthode M2 ;
- attribution d'un « + » à un indicateur à une seule composante.

Recommandation du Secrétariat : lorsqu'ils se heurtent à l'un des problèmes ci-dessus, **les évaluateurs d'une évaluation répétée ne doivent en aucun cas modifier explicitement une notation attribuée lors de l'évaluation antérieure.** En effet, cette notation a peut-être servi et des débats sans fin risquent d'en résulter. Toute erreur indiscutable de notation constatée dans une évaluation antérieure pourra faire l'objet d'explications dans la discussion sur les changements apportés aux notations attribuées aux indicateurs.

On peut insérer une note dans le tableau de présentation générale des indicateurs pour expliquer que les notations actuelle et précédente ne sont pas comparables, ou encore pour expliquer comment un point de vue différent sur la performance lors de l'évaluation précédente peut avoir depuis lors influencé les conclusions sur la direction des changements de performance du système de GFP.

Il importe aussi de noter que l'évaluation de chaque indicateur doit s'articuler avec l'évaluation précédente. En d'autres termes, il faut s'appliquer à expliquer la trajectoire de la performance en prenant pour point de départ l'évaluation précédente, et non évaluer l'indicateur isolément de l'évaluation précédente avant de comparer les notations. En effet, cette dernière méthode risque de ne pas prendre en compte le type de différences souligné plus haut.

Bonnes pratiques dans la conduite d'une évaluation répétée

Enfin, il peut être souhaitable, quoique pas toujours possible, de dire pourquoi la performance s'est améliorée – dans la mesure où il peut être alors nécessaire de procéder à une analyse plus détaillée que celle que le Rapport de performance de la GFP est supposé fournir, en creusant la question des causes sous-jacentes (comme par exemple les capacités disponibles).

7 Rapports sur les indicateurs individuels

Le Secrétariat recommande la présentation de ces rapports sous forme de tableau, comme suit :

Indicateur	Notation 2006	Notation 2010	Évolution de la performance	Autres facteurs
PI-1	C	B	La performance semble s'améliorer : 2006 : 6 %, 11 %, 18 % 2010 : 5 %, 11 %, 6 %	Il n'est pas sûr que tous les fonds des projets extérieurs aient été exclus des données de l'évaluation 2006, mais ce point est probablement mineur.
PI-4 (i)	A	C	L'évolution de la performance est incertaine, bien que les arriérés indiqués augmentent de 1 % en 2006 à 6 % en 2010.	L'évaluation 2006 utilise uniquement des données sur les ordres de paiement en attente, et non pas sur les factures en souffrance.

8 Conclusions sur l'évolution globale de la performance du système de GFP

L'augmentation du nombre d'évaluations répétées a suscité plusieurs tentatives d'établissement d'une mesure globale des évolutions de la performance. Cette mesure globale peut être particulièrement tentante lorsque certains indicateurs affichent des progrès tandis que d'autres font état d'une dégradation de la performance. L'agrégation implique convertir les notations ordinales de l'indicateur en notations numériques et de pondérer chacun des indicateurs. Néanmoins, comme il n'existe pas de méthode scientifiquement correcte pour effectuer ces conversions, le Secrétariat ne soutient ni ne recommande aucune méthode particulière. [Pour une analyse détaillée des problèmes méthodologiques posés par l'agrégation des scores PEFA, consulter le document *Issues in Comparison and Aggregation of PEFA Assessment Results Over Time and Across Countries* disponible sur le site du PEFA (www.pefa.org).]

Il ressort clairement de la présente note d'orientation que la méthode recommandée consiste à comparer les scores attribués à chaque indicateur et à mettre à profit la description narrative du Rapport de performance de la GFP pour tirer les conclusions rendant compte des différences de notation et ainsi de toute évolution de la performance du système de GFP. Le Résumé analytique de l'évaluation doit servir à tirer des conclusions globales sur la direction et l'ampleur des changements en soulignant l'importance relative des diverses évolutions des indicateurs et leurs liens.

9 Note de faisabilité / Termes de référence

Il faut accorder un soin particulier et précis au suivi des progrès retracé la note de faisabilité / les termes de référence. Pour que l'on puisse suivre correctement et comprendre l'évolution de la performance de la GFP, et se fier aux comparaisons de notations d'un indicateur au fil du temps, la description narrative du rapport de performance doit permettre de comprendre l'objet et l'ampleur de ce changement.

Les évaluateurs doivent donc être bien conscients de leur tâche : comparer les évaluations et le suivi de la performance dans le temps (et en rendre compte). La note de faisabilité / les termes de référence doivent exiger de l'équipe d'évaluation qu'elle explique tous les facteurs affectant l'évolution de la notation, indicateur par indicateur, qu'elle repère l'évolution de la performance et qu'elle s'assure que les lecteurs du document d'évaluation sont en mesure de suivre les changements par rapport à l'évaluation précédente. Cela peut représenter une partie substantielle de la rédaction du rapport, même si une précédente évaluation (de qualité) constituerait normalement un avantage pour le travail d'évaluation.

Pour que l'évaluation répétée permette de décrire une trajectoire de progrès exploitable, il faut articuler l'évaluation de chaque indicateur avec l'évaluation précédente, autrement dit s'employer à rendre compte de la trajectoire de performance en prenant pour point de départ l'évaluation précédente, et non évaluer l'indicateur isolément de l'évaluation précédente avant de comparer les notations (le danger de cette méthode est que la trajectoire de progrès peut pâtir de différences d'échantillonnage des définitions, de l'ajout de nouvelles informations, ou bien de différences d'interprétations ou de jugements dans les cas limites). La note de faisabilité / les termes de référence doivent être explicites sur ce point, comme le montre l'exemple de l'**ANNEXE A**.

Voici un exemple de note de faisabilité / termes de référence **susceptibles** de livrer une évaluation répétée satisfaisante :

- *« La mission d'évaluation a pour objectif de rédiger un rapport complet de performance de la gestion des finances publiques (RP-GFP), préparé selon la méthodologie PEFA, et ce, dans le but de fournir une analyse de la performance globale des systèmes de GFP du pays et d'assurer le suivi des progrès au regard des indicateurs PEFA depuis l'évaluation précédente. Cette analyse doit permettre de mesurer l'évolution de la performance dans le temps.*
- *Cette évaluation a pour objectifs spécifiques :*
 1. *de mettre à jour la présentation générale de la performance de la GFP, conformément au Cadre de mesure de la performance PEFA ;*
 2. *de déterminer et d'expliquer le degré d'amélioration de la performance basée sur les scores attribués aux indicateurs PEFA, en comparant avec les résultats constatés lors de l'évaluation précédente ;*
 3. *d'évaluer les résultats de l'examen PEFA, c'est-à-dire l'évolution de la performance en termes d'activités des projets et les conséquences possibles sur les notations obtenues.*
- *Les évaluateurs :*
 - a) *disposeront du rapport final de l'évaluation précédente*
 - b) *se verront communiquer tous les commentaires des pairs sur l'évaluation précédente*
 - c) *pourront s'entretenir avec le responsable de l'équipe d'évaluation précédente afin de se faire expliquer dans le détail, s'il y a lieu, les notations attribuées ».*

Bonnes pratiques dans la conduite d'une évaluation répétée

Voici un extrait de note de faisabilité / de termes de référence **non susceptible(s)** de déboucher sur une évaluation répétée satisfaisante :

- « *Il faut faire le point sur les progrès depuis l'exercice précédent* » (mais on n'attend rien de plus précis de la part des évaluateurs, pas même de consulter les scores précédents).

Annexe A

Exemple : Note de faisabilité / Termes de référence

Objectifs spécifiques

1. Mettre à jour la présentation générale de la performance de la GFP, conformément au Cadre de mesure de la performance PEFA.
2. Déterminer et expliquer le degré d'amélioration de la performance en fonction des notations des indicateurs PEFA, en comparant avec les résultats de l'évaluation précédente.
3. Évaluer les résultats de l'examen PEFA, c'est-à-dire l'évolution de la performance vue sous l'angle des activités composant le programme de réforme et les conséquences possibles sur les notations.
4. Prendre connaissance des raisons pouvant avoir contribué à l'évolution des notations, comme par exemple :
 - Les changements de définitions
 - Une meilleure disponibilité des informations ou un accès accru à celles-ci
 - Un échantillonnage et un regroupement différents des informations
 - Une approche différente des jugements professionnels
 - Les inexactitudes méthodologiques de notation dans l'évaluation précédente, comme par exemple l'attribution d'une notation « D » en l'absence d'informations suffisantes pour mettre une note.
5. Les consultants doivent s'assurer :
 - que tous les facteurs qui modifient une notation, indicateur par indicateur, sont expliqués
 - que l'on a cerné l'évolution de la performance
 - que toute personne prenant connaissance du document puisse suivre l'évolution depuis l'évaluation précédente.