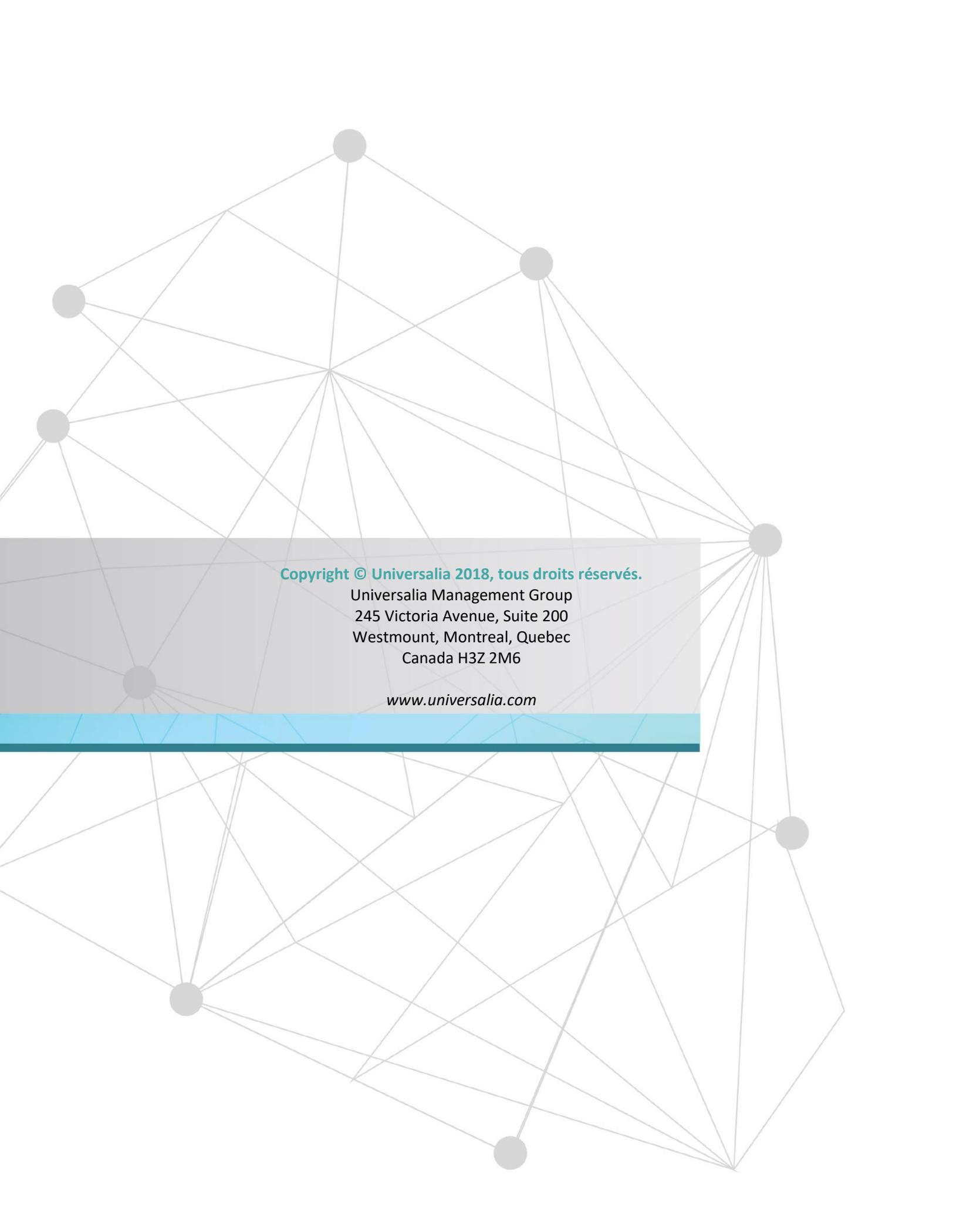


Appréciation de la qualité d'un échantillon de produits d'évaluation

Commandé par la Banque africaine de développement

RAPPORT FINAL | MAI 2018



Copyright © Universalia 2018, tous droits réservés.

Universalia Management Group
245 Victoria Avenue, Suite 200
Westmount, Montreal, Quebec
Canada H3Z 2M6

www.universalia.com

Remerciements

Universalialia tient à remercier l'équipe d'IDEV à Abidjan pour son esprit de coopération – notamment en mettant à disposition la documentation nécessaire –, son soutien logistique tout au long de cette mission et ses précieuses observations techniques et autres commentaires.

Acronymes

AQ	Appréciation de la qualité
BAD	Banque africaine de développement
ECG	Groupe de coopération pour l'évaluation
EdS	Égalité des sexes
GBR	Gestion basée sur les résultats
IDEV	Évaluation indépendante du développement
OCDE-CAD	Organisation de coopération et de développement économiques/Comité d'aide au développement
TdC	Théorie du changement
TdR	Termes de référence
UMG	Universal Management Group

Résumé analytique

Finalité globale, objectifs et portée de l'appréciation de la qualité

La présente analyse des produits d'évaluation d'IDEV avait pour but de tirer des enseignements susceptibles de contribuer à l'amélioration des produits en question. Elle visait notamment à déterminer si la qualité s'était améliorée depuis la mise en œuvre de la stratégie d'IDEV 2013-2017 en comparant, sur la base d'un échantillon, la qualité des évaluations réalisées entre 2009 et 2012 (avant la mise en œuvre de la stratégie) et celle des évaluations livrées vers la fin de sa mise en œuvre (2015-2017).

Au total, 22 évaluations ont été analysées, soit près de la moitié des 47 évaluations éligibles. L'échantillon était représentatif des principaux types d'évaluations menées par IDEV : évaluations de stratégie pays, évaluations thématiques et sectorielles, évaluations institutionnelles et de processus, mais aussi évaluations d'impact, évaluations régionales, évaluations groupées et synthèses d'évaluation. Autant que possible, des correspondances ont été établies entre les évaluations, sur la base de leur nature, entre la première période (neuf évaluations) et la seconde (treize évaluations).

Méthodologie

Cette appréciation de la qualité comportait une méta-analyse de divers produits d'évaluation. L'appréciation a été élaborée pour respecter les normes de qualité de l'OCDE-CAD relatives aux évaluations et aux bonnes pratiques d'évaluation normalisées du Groupe de coopération pour l'évaluation. Elle consistait essentiellement en une étude documentaire d'un échantillon de 22 évaluations, comprenant des rapports de synthèse et des documents associés, tels que des rapports techniques complets, des termes de référence, des documents d'orientation, des rapports de démarrage et toute la documentation pertinente disponible. Les constatations et les recommandations de l'appréciation de la qualité ont été présentées au personnel d'IDEV et affinées lors d'un atelier interactif qui s'est tenu à Abidjan le 27 mars 2018.

Un modèle standard comprenant 26 critères d'appréciation de la qualité, organisés en quatre groupes selon des normes internationales, a été utilisé pour évaluer et noter ces critères en se basant sur une échelle à quatre points. Les groupes incluent : le contexte, l'objet, la finalité, les objectifs et la portée de l'évaluation (groupe 1) ; la conception et la méthodologie (groupe 2) ; le processus (groupe 3) ; la clarté et la fiabilité des rapports (groupe 4). Les notes moyennes des groupes ont été calculées et codifiées par couleur, et les appréciations générales des évaluations ont été déterminées en fonction du pourcentage des notes obtenues pour chaque critère qui ont été jugées satisfaisantes ou très satisfaisantes (équivalant à 3 et 4).

Le processus d'appréciation de la qualité comportait trois étapes : l'appréciation individuelle de chaque évaluation par deux évaluateurs ; l'élaboration d'une seule appréciation pour chaque évaluation ; et l'harmonisation de toutes les appréciations par l'équipe entière (comprenant trois membres) pour

assurer la cohérence et la qualité de l'ensemble. Des techniques d'analyse quantitatives et qualitatives ont été utilisées pour préparer les résultats de l'appréciation de la qualité.

Certaines contraintes ont posé des limites à l'appréciation de la qualité : des données disponibles limitées pour certaines évaluations ; des données statistiques peu significatives en raison de la petite taille de l'échantillon ; et une comparabilité limitée entre les première et deuxième périodes pour certains types d'évaluations de l'échantillon qui n'ont pas été produites au cours de la première période.

Principales constatations

Pour l'ensemble des évaluations, **la majorité des critères d'appréciation de la qualité ont reçu la note « Satisfaisant » et « Très satisfaisant »**, avec une répartition assez homogène de ces notations dans les quatre groupes de critères.

La note la plus fréquente pour les quatre groupes de critères était « Satisfaisant », et le groupe de la conception et de la méthodologie montrait la plus grande variation de notes parmi toutes les évaluations.

On observe une nette amélioration de la qualité des évaluations au fil du temps de 2009-2012 à 2015-2017, les 13 évaluations de deuxième période recevant « Satisfaisant » ou « Très satisfaisant » comme note moyenne pour tous les groupes. Parmi les quatre groupes, le groupe 2 regroupant la conception et la méthodologie a enregistré la plus grande amélioration.

Parmi les sept critères liés au contexte, à l'objet, à la finalité, aux objectifs et à la portée de l'évaluation, la description de l'objet de l'évaluation et la portée de l'évaluation sont les critères les mieux notés, à l'opposé des questions transversales qui sont les moins bien notées dans les deux périodes considérées.

Au sein du groupe comprenant la conception et la méthodologie, la collecte de données, l'analyse et l'échantillonnage constituent le critère le mieux noté, puisqu'aucune évaluation n'a reçu une note « Très insatisfaisant » sur ce critère, **et l'alignement des questions d'évaluation avec les critères et les indicateurs dans une matrice d'évaluation est le critère le moins bien noté**, bien qu'on enregistre une amélioration significative de ce critère dans les évaluations de la deuxième période.

L'échelle de notation utilisée par les évaluations d'IDEV a également été jugée problématique. L'équipe d'évaluation de la qualité a pris acte de la récente décision d'IDEV de remplacer l'astreignante échelle de notation à 6 points par une échelle à 4 points pour les évaluations futures. Cependant, les notes de la nouvelle échelle (« Très insatisfaisant », « Insatisfaisant », « Satisfaisant », « Très satisfaisant ») restent problématiques dans une certaine mesure car elles ne permettent pas d'établir une évaluation plus nuancée entre « Insatisfaisant » et « Satisfaisant » là où cela serait pourtant pertinent ou tout du moins applicable.

Les liens entre les produits d'évaluation pour chaque évaluation étaient logiques dans l'ensemble, malgré un certain chevauchement des informations dans certains cas ou l'absence ponctuelle d'éléments clés dans le rapport de synthèse, probablement en raison des restrictions dans le nombre de pages autorisées pour ces rapports.

Bien qu'ils aient pu être effectivement suivis, les principes d'évaluation auxquels IDEV adhère, à savoir l'indépendance, l'impartialité, l'éthique, la coordination et l'alignement, le partenariat et le développement des capacités, n'ont pas été documentés de manière satisfaisante au cours des deux périodes. La participation des parties prenantes est un critère lié au processus qui a été légèrement moins bien noté dans la deuxième période : les parties prenantes clés de l'évaluation ainsi que leurs rôles et responsabilités n'étaient pas toujours systématiquement identifiés.

La grande majorité des rapports d'évaluation étaient bien structurés et comprenaient un résumé analytique concis, **présentaient des constatations fondées sur des données probantes, respectaient les critères et les questions d'évaluation et introduisaient des recommandations pertinentes. Cependant, ils n'identifiaient pas toujours explicitement les sources des données utilisées.**

Toutes les évaluations ont démontré des liens clairs entre les preuves, les constatations, les conclusions, les leçons apprises et les recommandations générales, malgré le manque de distinction parfois notable entre les éléments que chacune devait aborder. Dans près de la moitié des rapports examinés, les questions transversales n'ont pas bien été reprises dans les constatations et les conclusions, et ce dans les deux périodes, tandis que les éventuelles leçons apprises n'ont pas été incluses ou n'ont pas été correctement présentées en tant qu'enseignements dans la moitié des rapports, en particulier ceux de la deuxième période. Les recommandations, bien que globalement pertinentes, ne prévoyaient souvent pas de calendrier pour leur mise en œuvre.

Recommandations

1. Envisager l'ajout d'une cartographie des parties prenantes. Cette exigence devrait être stipulée dans les documents de planification initiale de l'évaluation, et dans la cartographie des parties prenantes élaborée par les évaluateurs et présentée dans les rapports techniques de démarrage et d'évaluation (en annexe si trop détaillée). De cette manière, toutes les parties prenantes essentielles (internes et externes) seraient identifiées et leurs rôles et contributions clairement décrits.
2. Demander à IDEV de définir les thèmes transversaux à hiérarchiser en se basant sur ce que la politique de la BAD a identifié comme important et spécifier les thèmes pertinents pour chaque type d'évaluation :
 - Les thèmes transversaux pertinents devraient être explicitement identifiés dès le départ (phase de démarrage) et efficacement intégrés de manière transversale, en commençant par la description du contexte et de l'objet de l'évaluation à travers les objectifs, la portée, les constatations, les conclusions et les recommandations de l'évaluation ;
 - Lorsque l'absence de données désagrégées sur ces thèmes empêche leur appréciation complète, l'évaluation peut être l'occasion d'approfondir ces questions en les intégrant dans le cadre d'évaluation et dans les instruments de collecte de données primaires correspondants, et de recommander à la direction de leur accorder plus d'attention dans le suivi des opérations de la Banque.
 - En ce qui concerne l'égalité des sexes en particulier, elle ne peut être considérée comme un thème transversal pertinent dans tous les types d'évaluations menées par IDEV, étant

donné que la BAD, comme toutes les institutions financières internationales, a une politique de promotion de l'égalité des sexes, IDEV devrait s'efforcer d'intégrer systématiquement les considérations de genre dans toutes les évaluations impliquant l'appréciation de la qualité de projets ainsi que dans d'autres types d'évaluations pour lesquelles il serait pertinent de le faire.

- Dans l'éventualité où il a besoin d'un soutien pour parvenir à une intégration efficace de ces questions, IDEV devrait demander une assistance technique spécialisée par le biais d'ateliers spécialisés, de webinaires, etc. visant à renforcer les capacités des chargés de projet dans ces domaines.
3. Veiller à ce que les principes d'évaluation de l'OCDE-CAD et de la BAD-IDEV, relatifs à l'indépendance, à l'impartialité, à l'éthique, au partenariat, à la coordination et au renforcement des capacités soient clairement et explicitement stipulés et expliqués dans les différents produits d'évaluation – de la conception au rapport final – et considérés comme faisant partie intégrante du processus d'évaluation. La preuve de la prise en compte de ces principes pourrait être fournie après la description de la méthodologie d'évaluation, dans l'introduction de divers outils de collecte de données et dans une clause d'exclusion de responsabilité dans les premières pages des rapports d'évaluation. Bien qu'une telle clause puisse être incluse dans les rapports de synthèse et les rapports techniques, des informations supplémentaires sur les principes d'évaluation pourraient être apportées dans l'annexe méthodologique des rapports de synthèse, compte tenu des restrictions relatives à la longueur de ces rapports.
 4. Envisager de renommer l'intitulé des notes de l'échelle de notation à quatre points pour créer une perspective qui reflète plus positivement un certain progrès vers une performance satisfaisante. Nous recommandons de remplacer les notes existantes par les termes « Insatisfaisant », « Partiellement insatisfaisant », « Satisfaisant » et « Très satisfaisant ».
 5. Identifier clairement les éléments de base que doit contenir chaque produit d'évaluation et clarifier les circonstances dans lesquelles les TdR, les documents d'orientation et les rapports de démarrage doivent être rédigés.
 6. Élaborer des lignes directrices pour les évaluateurs avec des définitions claires des termes clés de l'évaluation, pour bien distinguer le type d'information à inclure dans les différentes sections du rapport d'évaluation, par exemple les constatations de l'évaluation, les conclusions, les leçons apprises et les recommandations.
 7. S'assurer que les recommandations sont systématiquement débattues avec les sections opérationnelles pertinentes de la Banque et clairement classées par ordre de priorité ou selon un calendrier concordant pour aider à la prise de décision. Tout en évitant la formulation de recommandations trop prescriptives, IDEV devrait identifier les sujets de préoccupation dans les recommandations afin de les relier plus spécifiquement aux résultats des évaluations.
 8. Veiller à ce que les annexes du rapport de démarrage comprennent les outils de collecte de données et que les annexes du rapport de synthèse contiennent une matrice d'évaluation abrégée, c'est-à-dire un outil d'évaluation clé qui relie la portée de l'évaluation à la méthodologie de l'évaluation. La matrice abrégée pourrait inclure les critères d'évaluation et les principales questions correspondantes, ainsi qu'un échantillon d'indicateurs clés (à la fois

quantitatifs et qualitatifs, le cas échéant) et les méthodes de collecte de données pour chaque question principale.

Table des matières

RÉSUMÉ ANALYTIQUE	V
1 INTRODUCTION ET MISE EN CONTEXTE	1
1.1 Contenu du rapport	1
1.2 Stratégie d'évaluation indépendante de la BAD (IDEV) 2013-2017	1
2 FINALITÉ, OBJECTIFS ET PORTÉE DE LA MISSION	2
3 MÉTHODOLOGIE	3
3.1 Conception de méta-analyse et collecte de données	3
3.2 Échantillonnage, modèle et méthodes d'analyse	4
3.2.1 Échantillonnage	4
3.2.2 Modèle et critères d'appréciation de la qualité	5
3.2.3 Échelle de notation et justification	6
3.2.4 Analyse quantitative et qualitative	11
3.3 Limitations et actions d'atténuation	11
4 PRINCIPALES CONSTATATIONS	13
4.1 Analyse globale de la qualité des évaluations	13
4.1.1 Notations globales par groupe	13
4.1.2 Comparaison globale entre les types d'évaluations entre la première et la deuxième période	14
4.2 Analyse comparative des forces et des faiblesses des évaluations selon les critères entre la première et la deuxième période	18
4.2.1 Contexte, objet, finalité, objectifs et portée des évaluations	18
4.2.2 Conception et méthodologie	20
4.2.3 Processus	22
4.2.4 Clarté et fiabilité des rapports	24
5 CONCLUSIONS, LEÇONS APPRISSES ET RECOMMANDATIONS	28
5.1 Conclusions	28
5.2 Leçons apprises	30
5.3 Recommandations	30

Graphiques

Graphique 3.1	Répartition des évaluations par type et produits d'évaluation (n=22).....	5
Graphique 3.2	Processus d'appréciation de la qualité en trois étapes.....	10
Graphique 3.3	Répartition des évaluations parmi les évaluateurs.....	10
Graphique 4.1	Pourcentage des critères notés 3 et 4 par groupe, toutes évaluations confondues 13	
Graphique 4.2	Notes moyennes des groupes, toutes évaluations confondues (n=22).....	14
Graphique 4.3	Comparaison du pourcentage des critères notés 3 et 4 par type et période d'évaluation.....	15
Graphique 4.4	Comparaison des notes moyennes des groupes entre la première et la deuxième période, toutes évaluations confondues (première période n=9, deuxième période n=13).....	16
Graphique 4.5	Comparaison des notes moyennes des groupes par type d'évaluation, première et deuxième période.....	17
Graphique 4.6	Notations du groupe 1 : comparaison entre périodes (première période n=9, deuxième période n=13).....	19
Graphique 4.7	Notations du groupe 2 : comparaison entre périodes (première période n=9, deuxième période n=13).....	22
Graphique 4.8	Notations du groupe 3 : comparaison entre périodes (première période n=9, deuxième période n=13).....	24
Graphique 4.9	Notations du groupe 4 : comparaison entre périodes (première période n=9, deuxième période n=13).....	25

Tableaux

Tableau 3.1	Échelle de notation.....	6
Tableau 3.2	Exemples de quatre notations et appréciations différentes pour un même critère.....	8

Annexes

Annexe I	Liste des évaluations examinées.....	33
Annexe II	Modèle d'appréciation de la qualité.....	35
Annexe III	Pourcentage de critères jugés satisfaisants et très satisfaisants, toutes évaluations confondues.....	45

Annexe IV	Résultats de l'appréciation de la qualité par critère, toutes évaluations confondues.....	47
Annexe V	Résultats de l'appréciation de la qualité par critère, évaluations de la première période (2009-2012)	51
Annexe VI	Résultats de l'appréciation de la qualité par critère, évaluations de la deuxième période (2015-2017)	55
Annexe VII	Résultats de l'appréciation de la qualité par critère, évaluations par pays.....	57
Annexe VIII	Résultats de l'appréciation de la qualité par critère, évaluations institutionnelles.....	59
Annexe IX	Résultats de l'appréciation de la qualité par critère, évaluations thématiques et sectorielles	62
Annexe X	Résultats de l'appréciation de la qualité par critère, autres évaluations	65
Annexe XI	Résultats de l'appréciation de la qualité par groupe de critères, toutes évaluations confondues	69
Annexe XII	Résultats de l'appréciation de la qualité par groupe de critères, évaluations par pays	71
Annexe XIII	Résultats de l'appréciation de la qualité par groupe de critères, évaluations institutionnelles.....	73
Annexe XIV	Résultats de l'appréciation de la qualité par groupe de critères, évaluations thématiques et sectorielles.....	75
Annexe XV	Résultats de l'appréciation de la qualité par groupe de critères, autres évaluations	77
Annexe XVI	Résultats de l'appréciation de la qualité par groupe de critères et type d'évaluation.....	79
Annexe XVII	Résultats de l'appréciation de la qualité par groupe de critères, période d'évaluation et type d'évaluation	81
Annexe XVIII	Liste des documents consultés	83
Annexe XIX	Programme et rapport de l'atelier d'Abidjan.....	94

1 Introduction et mise en contexte

1.1 Contenu du rapport

Universal Management Group Limited (désormais « Universalia ») a le plaisir de soumettre ce rapport de synthèse final à IDEV, l'unité d'Évaluation indépendante du Développement de la Banque africaine de développement (BAD), portant sur l'appréciation de la qualité d'un échantillon de ses produits d'évaluation.

Cette introduction présente un aperçu du contexte de la mission, une brève description de la stratégie 2013-2017 de la BAD-IDEV et une évaluation de la mise en œuvre de la stratégie en cours, dont la présente analyse fait partie. Suite à cette introduction, le rapport comprend quatre sections principales :

- La section 2 donne un aperçu de la finalité, des objectifs et de la portée de la mission.
- La section 3 présente la méthodologie, et donc la conception, la collecte de données et les méthodes d'échantillonnage, la description du modèle d'appréciation de la qualité, les méthodes d'analyse, les limites et les mesures d'atténuation.
- La section 4 résume les principales constatations de l'appréciation de la qualité : après une analyse globale de la qualité des évaluations, elle présente une analyse comparative des forces et des faiblesses des évaluations par groupe et par critère pour chaque type d'évaluation (2009-2012 et 2015-2017).
- La section 5 présente les conclusions de l'appréciation de la qualité, les leçons apprises et les recommandations.

Certains exemples de cas sont mis en évidence dans des encadrés dans la section des constatations. Sont annexés au rapport : la liste des évaluations examinées ; le modèle d'appréciation de la qualité ; des données chiffrées détaillées ; la liste des documents consultés pour cette mission ; le programme et le rapport de l'atelier d'appréciation de la qualité, y compris la liste des participants.

1.2 Stratégie d'évaluation indépendante de la BAD (IDEV) 2013-2017

La première stratégie d'évaluation indépendante de la Banque a été élaborée dans le but de lier efficacement la politique d'évaluation indépendante de la Banque aux programmes d'IDEV sur une période de cinq ans entre 2013 et 2017. La politique « identifie trois objectifs complémentaires pour l'évaluation indépendante : l'apprentissage, la responsabilisation et la promotion d'une culture de

l'évaluation » dans le but de soutenir l'amélioration de l'efficacité du développement.¹ Dans la poursuite de ces trois objectifs, IDEV a réexaminé sa gamme de produits d'évaluation en diversifiant de plus en plus son portefeuille pour inclure davantage de groupes, de stratégies nationales et d'évaluations institutionnelles ainsi que des synthèses d'évaluation en introduisant de nouveaux produits tels que les évaluations de stratégies d'intégration régionale et les évaluations d'impact. IDEV a maintenu sa production traditionnelle d'évaluations thématiques et sectorielles, mais a réduit le nombre d'évaluations de performance de projets.²

Cette appréciation de la qualité des produits d'évaluation d'IDEV réalisée par Universalialia fait partie d'un examen complet de la mise en œuvre par IDEV de la stratégie actuellement menée en interne par IDEV (c'est-à-dire par auto-évaluation). IDEV a demandé à une compagnie externe de mener l'appréciation de la qualité pour améliorer la transparence, éviter les conflits d'intérêts potentiels et apporter une expertise supplémentaire à la mission. Les résultats de cette appréciation alimentent l'examen plus large de la mise en œuvre de la stratégie d'évaluation indépendante et aident à établir les nouvelles orientations stratégiques de la fonction d'évaluation indépendante de la BAD, en remplaçant ou en prolongeant la stratégie en 2018.

2 Finalité, objectifs et portée de la mission

Bien que cette appréciation de la qualité permette de rendre compte de l'amélioration de la qualité de l'évaluation, sa finalité principale était de tirer les enseignements qui aideront IDEV à continuer d'améliorer la qualité de ses évaluations.

Cette appréciation de la qualité visait en particulier à déterminer si la qualité des évaluations fournies par IDEV s'était améliorée depuis le début de la mise en œuvre de la stratégie. Par conséquent, l'appréciation visait à comparer la qualité des évaluations réalisées avant la période de mise en œuvre de la stratégie (c'est-à-dire 2009-2012) avec les évaluations livrées vers la fin de sa mise en œuvre (2015-2017).

La portée de cette mission, telle que définie dans les TdR et la proposition technique d'Universalialia, a été affinée au cours de la phase de démarrage. Elle est circonscrite par les éléments clés suivants :

- **Objet de l'appréciation de la qualité :** comme convenu avec IDEV, l'échantillon total comprend 22 évaluations (voir l'annexe I pour avoir la liste complète des évaluations examinées), ce qui représente près de 50 % du nombre total d'évaluation qui est de 47 (voir la totalité des évaluations ci-dessous). L'appréciation n'incluait pas les évaluations conjointes qui n'étaient pas menées par IDEV et ne couvrait pas d'autres produits d'évaluation connexes tels que les mémoires, les magazines et les événements.

¹ IDEV, *Stratégie d'évaluation indépendante de la Banque africaine de développement pour 2013-2017*, février 2013, p. 17.

² Ibid., p. 19. Bien que les résultats des évaluations de projets aient diminué, il convient de noter que chaque évaluation groupée comprend l'évaluation de plusieurs projets.

- **Typologie des évaluations :** l'échantillon est représentatif des principaux types d'évaluations menées par IDEV, telles que définies dans sa stratégie et son rapport de démarrage pour évaluer la mise en œuvre de la stratégie³, c'est-à-dire les évaluations de stratégies de pays, les évaluations thématiques et sectorielles, les évaluations d'entreprises et de processus, et d'autres telles que les évaluations d'impact, les évaluations régionales, les évaluations de groupes et les synthèses d'évaluation. Elle n'incluait pas les rapports individuels d'évaluation de la performance des projets (REPP).
- **Période examinée :** conformément aux termes de référence et au rapport de démarrage pour évaluer la stratégie d'IDEV, l'échantillon d'évaluations réalisées avant le début de la mise en œuvre de la stratégie (2009-2012) correspond aux « évaluations de la première période » tandis que les évaluations achevées vers la fin de la stratégie (2015-2017) sont appelées « évaluations de la deuxième période » dans ce rapport.
- **Nombre total d'évaluations :** le nombre d'évaluations admissibles pour la période pré-stratégie/début de mise en œuvre (c.-à-d. 2009-2013⁴) s'élevait à 20, contre 28 pour la période 2015-2017, dont une qui n'a finalement pas été prise en compte⁵ pour éviter tout conflit d'intérêt, puisque réalisée par Universalialia. Par conséquent, le nombre total d'évaluations admissibles était de 47.
- **Dimensions à couvrir :** dans le but de mener un examen aussi complet que possible de chaque évaluation, l'appréciation de la qualité comptait deux dimensions : i) la première portait sur les informations contenues dans le rapport de synthèse de l'évaluation ; et ii) la seconde consistait à examiner d'autres produits associés pour obtenir des informations supplémentaires sur divers aspects des critères d'évaluation qui ne figuraient pas dans le rapport de synthèse, avec par exemple, des détails sur la conception et la méthodologie, des informations complémentaires dans les annexes, etc. IDEV limite le nombre de pages de rapports de synthèse à 20 au maximum et les annexes des rapports à 20 pages également.

3 Méthodologie

3.1 Conception de la méta-analyse et collecte de données

Cette mission d'appréciation de la qualité comportait une méta-analyse de divers produits d'évaluation et était guidée par les normes de qualité de l'OCDE-CAD relatives aux évaluations⁶ et par les bonnes

³ IDEV, *Évaluation de la mise en œuvre de la Stratégie d'évaluation indépendante de la BAD 2013-17*, Rapport de démarrage V3, novembre 2017.

⁴ Alors que les évaluations réalisées en 2013 faisaient partie de la liste des évaluations admissibles, l'équipe d'appréciation de la qualité a choisi de les exclure de l'échantillon car elles coïncidaient avec le début de la période de mise en œuvre de la stratégie et ne représentaient donc pas une référence antérieure propice à la comparaison des périodes.

⁵ Banque africaine de développement, *Maroc : Évaluation de la stratégie et du programme de la Banque 2004-2014*, Rapport de synthèse, Évaluation de la stratégie pays d'IDEV, septembre 2016.

⁶ Comité d'aide au développement de l'OCDE, *Lignes directrices et publications du CAD*, « Normes de qualité pour l'évaluation du développement », 2010.

pratiques du Groupe de coopération pour l'évaluation⁷. Elle consistait essentiellement en une étude documentaire d'un échantillon de 22 évaluations, comprenant des rapports de synthèse et des documents associés, tels que des rapports techniques complets, des termes de référence, des documents d'orientation, des rapports de démarrage et toute la documentation pertinente disponible. Les constatations et les recommandations de l'appréciation de la qualité ont été présentées au personnel d'IDEV et affinées lors d'un atelier interactif qui s'est tenu à Abidjan le 27 mars 2018.

Les premiers échanges entre le personnel d'IDEV et celui d'Universalialia ont permis de parvenir à une compréhension commune de la mission. Il était important de noter que ce travail d'appréciation de la qualité était avant tout un exercice d'apprentissage constructif pour IDEV et ne visait pas les évaluations supervisées par les chargés de projet ayant obtenu le score le plus élevé ou le moins élevé, bien que des informations sur les évaluations les plus performantes aient été partagées avec la direction d'IDEV pour servir d'exemples de bonnes pratiques. Dans cet esprit, l'analyse des notations présentée dans ce rapport ne se réfère pas à des évaluations spécifiques.

Les échanges initiaux ont également été l'occasion de finaliser l'échantillon et d'affiner la méthodologie et l'outil d'appréciation de la qualité correspondant qu'Universalialia a ensuite présenté dans un rapport de démarrage. Lors du choix de l'échantillon, l'équipe d'appréciation de la qualité a tenté de faire correspondre les types d'évaluations des deux périodes considérées pour obtenir un échantillon représentatif permettant une comparaison dans le temps.

3.2 Échantillonnage, modèle et méthodes d'analyse

3.2.1 Échantillonnage

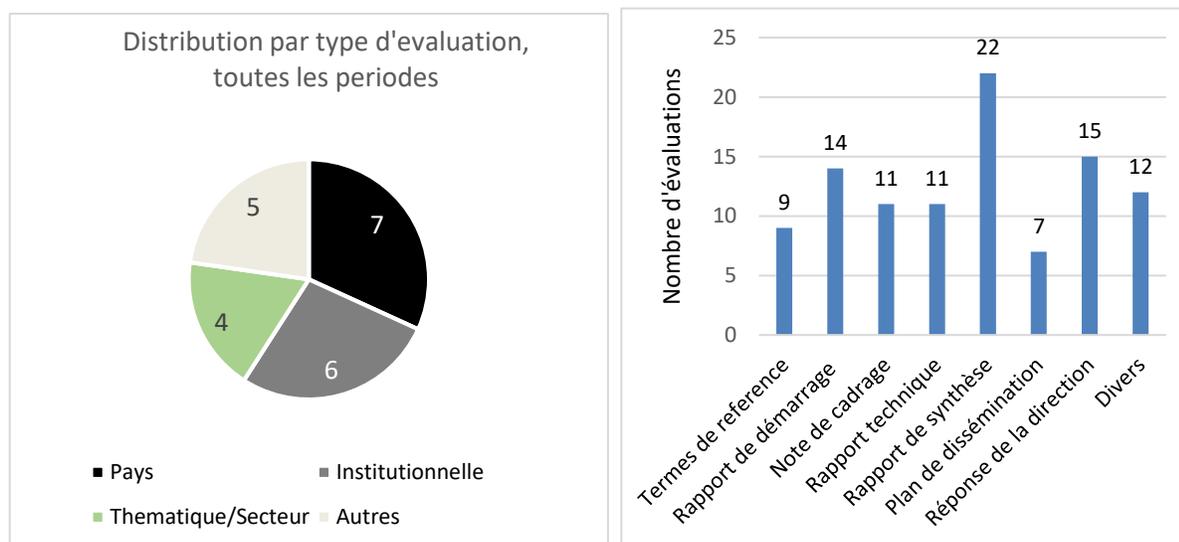
Lors du choix de l'échantillon, l'équipe d'appréciation de la qualité a tenté de faire correspondre les types d'évaluations des deux périodes considérées pour obtenir un échantillon représentatif permettant une comparaison dans le temps. L'équipe a également tenté d'obtenir une représentation proportionnelle des évaluations admissibles pour chaque période. Parmi les évaluations retenues, on compte 9 évaluations réalisées entre 2009 et 2011, représentant 45 % des évaluations éligibles pour la première période (2009-2013) et 13 évaluations de la deuxième période (2015 -2017), représentant 48 % des évaluations éligibles. D'autres facteurs ont influencé la sélection des évaluations représentant chaque période : classer par ordre de priorité les évaluations les plus récentes et les plus anciennes de chaque période où les changements dans le temps seraient les plus perceptibles ; tout en tentant de faire correspondre les types d'évaluations sélectionnés pour chaque période, inclure davantage d'évaluations dans la catégorie « Autres » de la deuxième période, car un nombre important d'évaluations dans cette catégorie ont été produites lors de la deuxième période comparativement à la première période ; et intégrer des évaluations représentant les pays anglophones et francophones.

Ainsi, conscients que les évaluations publiées en 2015 ont effectivement débuté en 2014, lorsque la stratégie d'évaluation d'IDEV était dans sa première phase de mise en œuvre, l'équipe a pris soin de

⁷Groupe de coopération pour l'évaluation, *Grand livre des normes de bonnes pratiques* en matière d'évaluation, novembre 2012 ; Groupe de coopération pour l'évaluation, *Normes de bonnes pratiques - Évaluation de la stratégie et du programme pays*, 2008.

sélectionner les évaluations de cette année uniquement lorsque cela s'avérait nécessaire, notamment quand le portefeuille manquait d'évaluations pour les années suivantes. Seules deux évaluations de 2015 ont été sélectionnées. Comme la plupart des 47 évaluations admissibles provenaient des catégories « Institutionnelles » (n=11) et « Pays » (n=17), six évaluations ont donc été sélectionnées dans la première catégorie et sept dans la deuxième. Bien que dans l'ensemble, il y avait également un nombre important d'évaluations éligibles dans la catégorie « Thématiques/Sectorielles » (n=10), quatre évaluations seulement ont été sélectionnées dans cette catégorie car peu d'entre elles avaient été réalisées sur la période 2015-2017, ce qui limitait les possibilités de comparaison. Enfin, cinq évaluations ont été sélectionnées dans la catégorie « Autres », représentant quatre types différents d'évaluations (synthèse, régionale, groupée et impact). Trois évaluations récentes dans cette catégorie n'étaient pas comparables avec les évaluations de la première période, mais comme toutes les autres évaluations, elles ont été évaluées principalement par rapport aux références internationales. Le graphique 3.1 ci-dessous montre la répartition par type des évaluations dans l'échantillon et inclut le nombre d'évaluations pour lesquelles chaque produit d'évaluation a été trouvé. La catégorie « Divers » comprend différents documents tels que des études de cas, des notes d'orientation/de concept, des rapports d'étape, des documents d'information, des notes explicatives, un manuel, des notes de service et diverses appréciations des résultats de projets pour les évaluations pays.

Graphique 3.1 Répartition des évaluations par type et produits d'évaluation (n=22)



3.2.2 Modèle et critères d'appréciation de la qualité

L'appréciation de la qualité a impliqué l'utilisation d'un modèle standard développé et convenu avec l'équipe IDEV pendant la phase de démarrage. Le modèle comprend 26 critères d'appréciation de la qualité, dont les éléments sont définis pour chacun conformément aux normes d'évaluation internationales, à savoir les normes de qualité de l'OCDE-CAD et les lignes directrices du Groupe de coopération pour l'évaluation (GCE) (voir modèle en annexe II).⁸

⁸ Le modèle a d'abord été développé dans Word, tel que présenté à l'annexe II, puis converti en Excel pour faciliter la saisie des données et leur compilation et analyse ultérieurement.

Le modèle est divisé en trois parties. La première partie est essentiellement descriptive, indiquant des informations de base sur l'évaluation : titre, type, année de livraison, documents examinés. La deuxième partie est divisée en quatre sections représentant les groupes utilisés pour structurer l'analyse selon une série de critères d'appréciation de la qualité :

- i) **Groupe 1** : sept critères liés au contexte, à la finalité, au but, aux objectifs et à la portée de l'évaluation ;
- ii) **Groupe 2** : quatre critères liés à la conception et à la méthodologie de l'évaluation ;
- iii) **Groupe 3** : cinq critères liés au processus d'évaluation ; et
- iv) **Groupe 4** : dix critères liés à la clarté et à la fiabilité des rapports.

Chaque critère est numéroté et comprend une explication en italique indiquant les éléments particuliers que les évaluateurs devraient rechercher lors de leur travail d'appréciation.

Les critères intègrent à la fois des dimensions quantitatives et qualitatives, c'est-à-dire que les évaluateurs déterminent combien d'éléments identifiés ont été trouvés pour chaque critère dans l'évaluation et examinent, entre autres critères, la clarté, la pertinence et la cohérence des éléments trouvés. Chaque critère a été évalué à l'aide d'une échelle à quatre points décrite ci-dessous et définie au tableau 3.1. Chaque section fournit ensuite une note moyenne et une évaluation équilibrée de la qualité en fonction des notes des critères individuels pour ce groupe. La troisième partie du modèle représente une appréciation globale de l'évaluation afin de résumer ses principales forces et faiblesses et d'inclure un pourcentage de tous les critères notés « Satisfaisant » et « Très satisfaisant » parmi les groupes.

3.2.3 Échelle de notation et justification

Les définitions pour chaque notation sont fournies dans le tableau 3.1 ci-dessous. Pour les quelques cas qui présentaient un critère ou certains éléments d'un critère non applicables, par exemple, pour certaines évaluations de la catégorie « Autres », la mention « Non applicable » leur a été attribuée et ils ont été exclus de la notation. La note globale pour chacun des quatre groupes évalués reflète une moyenne des notes attribuées aux critères individuels correspondants selon la même échelle à 4 points, et s'est vu attribuer un code couleur⁹. Étant donné que tous les critères sont pondérés de manière égale, les notes globales des groupes ont été calculées en tant que moyenne de toutes les notes individuelles, c'est-à-dire que tous les points ont été additionnés et divisés par le nombre de critères de la section. Les scores comportant plusieurs chiffres ont été arrondis à la première décimale inférieure ou supérieure et le code de couleur correspondant équivaut à la note numérique la plus proche. Cependant, les appréciations globales d'évaluation dans la troisième partie du modèle reflètent un pourcentage de critères individuels notés « Satisfaisant » ou « Très satisfaisant » (notes de 3 et 4) et les commentaires correspondants résument les principales forces et faiblesses de l'évaluation. Les pourcentages ont été arrondis à la première décimale inférieure ou supérieure.

Tableau 3.1 Échelle de notation

⁹ 1 = rouge ; 2 = jaune ; 3 = vert clair ; 4 = vert foncé.

NOTE (CRITÈRE)	SIGNIFICATION		NOTE (GROUPE)
4	Très satisfaisant	L'évaluation respecte tous ou presque tous les éléments du critère d'appréciation de la qualité et présente des lacunes négligeables voire inexistantes.	3,5 – 4
3	Satisfaisant	L'évaluation répond à la majorité des éléments du critère d'appréciation de la qualité et présente quelques lacunes.	2,5 – 3,49
2	Insatisfaisant	L'évaluation ne respecte que quelques-uns des éléments du critère d'appréciation de la qualité et présente des lacunes importantes.	1,5 – 2,49
1	Très insatisfaisant	L'évaluation ne répond pratiquement à aucun des éléments du critère d'appréciation de la qualité et présente de graves lacunes.	1 – 1,49
—	Non applicable	Le critère ou certains éléments du critère ne sont pas applicables.	Non noté

Des commentaires rédigés par les évaluateurs accompagnent toutes les notations, fournissant une justification pour l'appréciation de chaque critère et pour les notations globales des groupes, et permettant une comparaison claire des forces et des faiblesses des évaluations entre les deux périodes à l'étude. Tous les membres de l'équipe sont arrivés à une compréhension commune des critères d'appréciation de la qualité pour s'assurer que tous adoptent la même approche. Des éléments quantitatifs et qualitatifs, intégrés aux critères, sont à prendre en compte pour déterminer une note pour chaque critère, tout en tenant compte des documents, comme par exemple un rapport de synthèse, un rapport technique, un rapport de démarrage, etc., dans lesquels l'information a été trouvée ou non. Un effort a également été fait pour éviter de pénaliser le même élément deux fois ou plus. Par exemple, si aucune finalité n'est définie pour une évaluation (critère 3) et si ce critère est mal noté, d'autres critères qui se rapportent à la « cohérence avec la finalité », par exemple, les critères 4, 5 et 6, ne sont pas correctement notés non plus puisque la cohérence avec la finalité ne peut pas être vérifiée. Au lieu de cela, ces critères sont évalués en termes de cohérence interne et de clarté, par exemple le critère 4, et/ou la cohérence avec les objectifs de l'évaluation, par exemple les critères 5 et 6.

En outre, puisque les critères d'appréciation de la qualité s'inspirent des normes internationales établies en la matière, lorsque les informations jugées importantes par les évaluateurs car susceptibles d'apporter une compréhension élémentaire et adéquate de l'évaluation, et, par souci de conformité avec les normes internationales, ne figurent pas dans le rapport de synthèse mais se trouvent dans le rapport technique complet, les annexes au rapport de synthèse, ou seulement dans le résumé analytique, les notes sont ramenées de « Très satisfaisant » à « Satisfaisant » (notes 4 et 3 respectivement), étant donné que ces informations auraient dû apparaître dans le texte principal du rapport de synthèse.¹⁰

Le tableau suivant fournit des exemples de notes allant de 1 à 4 pour un critère sélectionné afin d'illustrer comment les notes ont été justifiées dans quelques cas. Il convient de noter que ces exemples, tirés des appréciations de la qualité, doivent simplement servir d'illustration et ne sont pas destinés à déterminer pleinement le type d'informations qui doivent nécessairement apparaître pour un critère particulier et, de toute façon, certains commentaires peuvent ne pas être pertinents en fonction du type

¹⁰ On ne trouve que quelques exemples de cela dans l'appréciation de la qualité.

d'évaluation. Par exemple, certains commentaires ne s'appliquent pas aux évaluations institutionnelles. Comme mentionné au début de cette section, lorsqu'un critère ou certains éléments d'un critère sont jugés non applicables, cela n'affecte pas la notation.

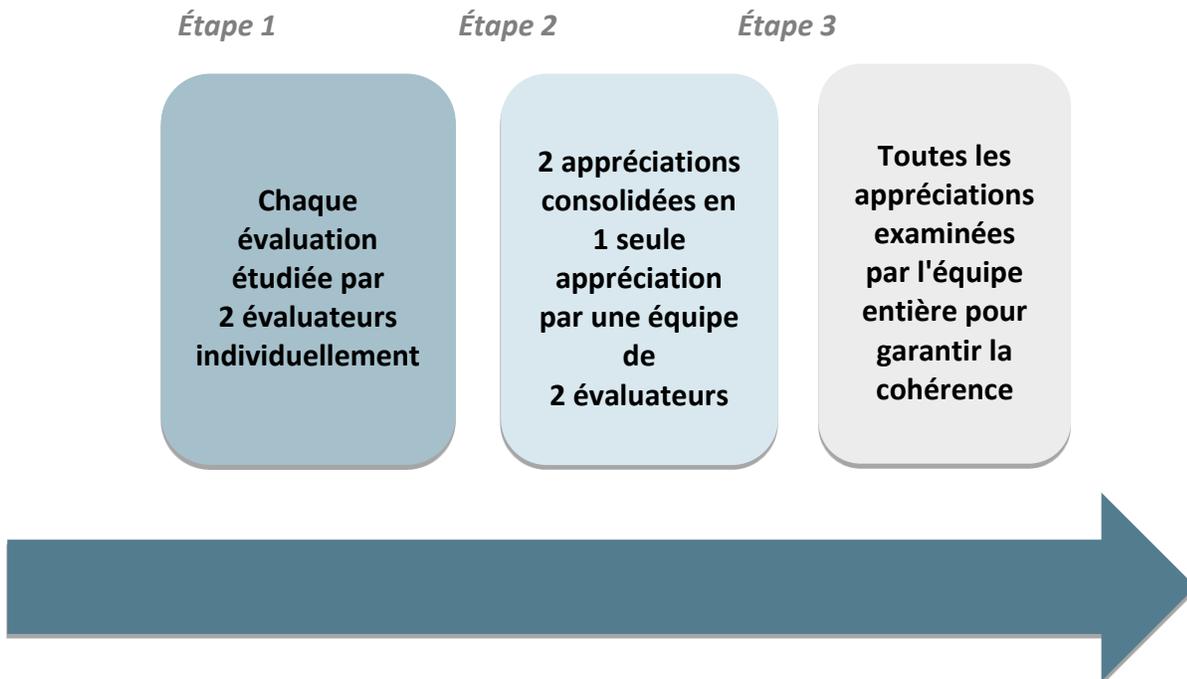
Tableau 3.2 Exemples de quatre notations et appréciations différentes pour un même critère

CRITÈRE	NOTATION/APPRÉCIATION			
	1	2	3	4
<p><u>Groupe 1</u> 1. L'historique/le contexte de l'objet de l'évaluation est clairement décrit et inclut tous les facteurs pertinents. <i>Principaux facteurs sociaux, culturels, politiques, économiques et institutionnels qui sont pertinents pour l'objet de l'évaluation.</i></p>	<p>L'évaluation ne comprend pas de section qui contextualiserait l'aide de la Banque au cours de la période considérée. Il serait important d'inclure une section sur le contexte, présentant peut-être des informations générales sur les principales tendances sociales, culturelles, politiques et économiques en Afrique au cours de cette période. Cela aiderait le lecteur à mieux comprendre dans quel contexte l'aide de la Banque a été fournie.</p>	<p>Bien que le rapport de démarrage contienne des informations contextuelles pertinentes qui justifient l'évaluation, le rapport de synthèse fournit très peu d'informations sur le contexte, avec des informations strictement axées sur le contexte institutionnel. L'historique du rapport technique présente l'évaluation mais n'inclut pas de description du contexte. Certains éléments contextuels de base décrits dans le rapport de démarrage doivent être inclus dans le rapport de synthèse.</p>	<p>Le rapport de synthèse comprend une section sur le contexte du pays, qui fournit un aperçu détaillé du contexte géographique et politique, ainsi qu'un aperçu des stratégies nationales de développement, de la croissance économique du pays et des facteurs institutionnels qui entravent le développement économique. Il comprend une brève description des programmes nationaux successifs qui ont servi de base aux stratégies et programmes de la Banque au cours de la période considérée. Des informations sur le contexte social et culturel du pays, y compris la situation de groupes ethniques et marginalisés, pourraient être fournies dans le texte principal du rapport de synthèse car il s'agit d'informations pertinentes dans le contexte de cette évaluation, bien qu'il y ait des indicateurs socio-économiques</p>	<p>L'évaluation fournit des informations pertinentes sur le contexte de la BAD, sur le continent en général et dans certains pays concernés, ainsi que sur l'évolution de la situation au cours de la période couverte par l'évaluation. L'évaluation examine le contexte dans lequel les initiatives ont été mises en œuvre, y compris les changements récents dans le contexte international et régional, les développements dans l'architecture de l'aide internationale et l'environnement économique international. Cette section traite également des facteurs politiques, comme les conflits et l'instabilité (par exemple, en Sierra Leone, en RDC) ou les crises politiques (par exemple, en Côte d'Ivoire, au Zimbabwe). La section traite également de la façon dont l'accent mis sur l'efficacité de l'aide (menant à la Déclaration de Paris) a affecté l'objet de</p>

CRITÈRE	NOTATION/APPRÉCIATION			
	1	2	3	4
			inclus dans les annexes.	l'évaluation. Enfin, cette partie présente le cadre institutionnel de la Banque (structures et politiques), dans lequel l'objet a évolué.

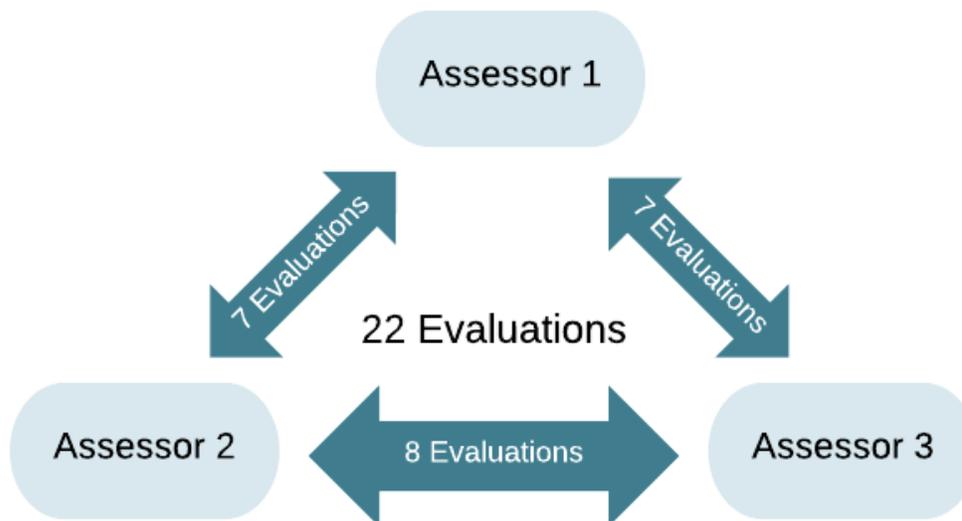
Le processus d'appréciation de la qualité comportait trois étapes, comme illustré dans le graphique 3.2 ci-dessous. Premièrement, chaque évaluation a été examinée en parallèle par deux évaluateurs. Deuxièmement, les deux évaluateurs de chaque évaluation ont discuté de leurs notes/commentaires et les ont regroupés en une seule appréciation pour chaque évaluation. Troisièmement, et une fois toutes les évaluations consolidées, l'équipe entière a examiné toutes les appréciations pour assurer la cohérence et la qualité de l'ensemble des travaux, apportant des modifications mineures à certaines notations afin d'aboutir à des appréciations harmonisées pour tous les critères.

Graphique 3.2 Processus d'appréciation de la qualité en trois étapes



La figure 3.3 ci-dessous illustre la répartition des évaluations entre les évaluateurs. Les évaluateurs 1 et 2 ont réalisé leurs appréciations de la qualité individuellement et ont ensuite consolidé leurs appréciations de sept évaluations, tandis que les évaluateurs 1 et 3 ont réalisé leurs appréciations individuellement pour ensuite consolider leurs appréciations de sept autres évaluations, et les évaluateurs 2 et 3 ont réalisé leurs appréciations individuellement pour ensuite consolider leurs appréciations des huit évaluations restantes pour un total de 22 évaluations. Ainsi, un évaluateur a examiné au total 14 évaluations tandis que deux évaluateurs ont examiné au total 15 évaluations. Les trois évaluateurs ont ensuite discuté des notes de chacune des 22 évaluations afin d'assurer la cohérence globale.

Graphique 3.3 Répartition des évaluations parmi les évaluateurs



3.2.4 Analyse quantitative et qualitative

À la fin de la troisième étape, l'équipe d'examen a utilisé des méthodes d'analyse de données mixtes pour obtenir les résultats de l'appréciation de la qualité, comme suit :

- **Analyse descriptive qualitative** pour résumer les principales caractéristiques des évaluations (par exemple, le type et l'année de l'évaluation, le type de documents disponibles et examinés pour chaque évaluation) ;
- **Analyse descriptive quantitative** pour déterminer le nombre d'évaluations associées aux principales caractéristiques identifiées ;
- **Analyse comparative qualitative et quantitative** pour comparer les constatations (notations et appréciations) pour chacun des 26 critères d'appréciation de la qualité dans toutes les évaluations, et plus particulièrement entre les évaluations de la première période et celles de la deuxième période au sein et entre les différents types d'évaluations.
- **Analyse groupée** pour comparer les appréciations des critères au sein et entre les quatre groupes de critères (groupes) pour toutes les évaluations, et plus particulièrement entre les évaluations de la première période et celles de la deuxième, et entre les différents types d'évaluations ;
- **L'analyse de contenu** a été utilisé pour identifier les tendances, les thèmes et les modèles communs qui ont émergé pour chacune des unités clés d'analyse ou, inversement, pour mettre en évidence des tendances opposées.

En résumé, l'analyse visait à identifier les forces et les faiblesses des évaluations, les thèmes émergents et les changements observés au fil du temps.

3.3 Limitations et actions d'atténuation

- **Données disponibles limitées pour certaines évaluations.** L'étude documentaire était nécessairement limitée par la quantité d'informations disponibles qui variait d'une évaluation à l'autre. Alors qu'IDEV fournissait tous les documents qui étaient à sa disposition pour chaque évaluation du portefeuille, la liste complète des documents n'a pas pu être rassemblée pour toutes les évaluations, en particulier celles de la première période pour lesquelles moins de documents étaient disponibles. Il est possible que certains documents liés à des évaluations de la première période aient été perdus ou que ces évaluations aient été moins documentées. Comme toutes les évaluations comportaient au moins le rapport de synthèse, ces absences ont été atténuées en mettant ce rapport au centre de l'évaluation de la qualité. Ce n'est que lorsque certains critères impliquaient nécessairement une comparaison d'informations des rapports de synthèse avec d'autres sources, ou lorsqu'il y avait des lacunes dans les rapports de synthèse empêchant une évaluation adéquate de certains critères, que les évaluateurs examinaient d'autres documents disponibles pour combler ces insuffisances.
- **La nature « statique » de l'étude documentaire peut ne pas tenir compte des évolutions en cours.** De nouvelles politiques ou orientations en matière de programmation et d'évaluation peuvent être en place ou en cours de développement par la BAD-IDEV puisque la dernière évaluation de l'échantillon a été réalisée en 2017 et sort donc du cadre de cette appréciation de la qualité. Cette appréciation fournit un regard rétrospectif sur les différences de qualité des

évaluations entre deux périodes spécifiques. Bien qu'elle ne tienne pas compte des plus récents changements introduits dans le processus pratique d'évaluation, elle contient néanmoins des informations qui peuvent éclairer les initiatives actuelles et futures. De plus, lorsque des changements ou des clarifications dans le processus pratique d'évaluation ont été portés à l'attention du représentant de l'équipe d'appréciation de la qualité à l'atelier d'apprentissage à Abidjan, ces derniers ont été consignés dans le rapport.

- **Signification statistique limitée.** Bien que l'échantillon soit représentatif des divers types d'évaluations menées par IDEV, en raison de la petite taille de l'échantillon, les résultats de cette évaluation n'étaient pas censés être statistiquement significatifs.
- **Difficultés inhérentes à la réconciliation d'éléments quantitatifs et qualitatifs au sein d'une notation.** Les critères d'appréciation comprenaient des éléments quantitatifs et qualitatifs qui devaient être équilibrés afin de déterminer une note juste pour chaque critère. Cela a été facilité par l'identification, dans le modèle, de divers éléments que les évaluateurs devraient considérer pour chaque critère à la fois quantitativement (nombre d'éléments) et qualitativement (clarté, pertinence, cohérence, etc.). Ces éléments ont ensuite été complétés par des indices de notation plus spécifiques aux évaluateurs pour chaque critère.
- **Les défis posés par la nature atypique de certaines évaluations,** plus précisément celles de la catégorie « Autres », comme les évaluations groupées ou les synthèses d'évaluation, ont rendu plus difficile la notation des critères d'appréciation de la qualité élaborés à partir de critères d'évaluation standard plus adaptés aux évaluations formatives ou sommatives d'interventions spécifiques. Lorsque certains critères ne s'appliquaient pas clairement à ces évaluations, une note « Non applicable » a été donnée, celle-ci n'ayant aucun effet négatif sur les notes moyennes des groupes.
- **Comparabilité limitée entre la première et la deuxième période pour certains types d'évaluation de la catégorie « Autres »** du fait que ces types d'évaluation n'étaient pas réalisés au cours de la première période. Cela ne constituait pas une difficulté importante puisque la qualité de tous les produits d'évaluation a été appréciée en fonction des mêmes critères, conformément aux références internationales.
- Toute **tendance potentielle de l'évaluateur** à interpréter des critères de notation et de l'information contenue dans les documents d'évaluation a été évitée grâce à des consultations collégiales au début du projet afin d'assurer une compréhension commune des critères d'appréciation de la qualité, qui elles-mêmes ont été suivies d'un processus rigoureux d'analyse des données en trois étapes impliquant deux évaluateurs examinant individuellement chaque évaluation, les réunissant ensuite pour consolider leurs notations, et amenant l'équipe entière d'évaluateurs pour se concerter et harmoniser les notes pour chaque critère dans les 22 appréciations de qualité. De plus, après la réalisation d'un certain nombre d'évaluations par chaque binôme, des indices de notation ont été ajoutés à l'attention des évaluateurs à côté de chaque critère du modèle pour traiter des cas particuliers.
- **L'intitulé et les définitions relatives à l'échelle de notation à quatre points ne permettaient pas d'attribuer une note entre « Insatisfaisant » et « Satisfaisant »** et donc une appréciation plus nuancée de certains critères pour lesquels l'attribution d'une note intermédiaire paraissait plus appropriée. Pour résoudre cela, les évaluateurs se sont appuyés sur leur jugement pour décider si l'information à l'étude méritait plus l'une ou l'autre de ces notes, c.-à-d. s'ils penchaient davantage vers une appréciation insatisfaisante ou satisfaisante.

4 Principales constatations

4.1 Analyse globale de la qualité des évaluations

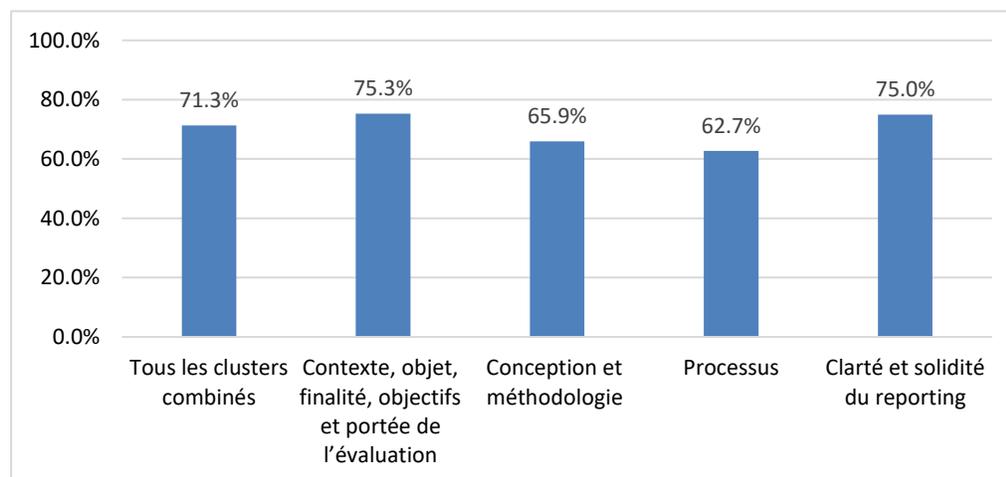
4.1.1 Notations globales par groupe

Étant donné qu'aucune note globale moyenne n'a été calculée pour chaque évaluation, l'appréciation globale de la qualité des évaluations se reflète dans le pourcentage de notes « Satisfaisant » et « Très satisfaisant » de chaque évaluation, ainsi que dans la note moyenne déterminée pour chaque groupe de critères, tel que défini dans la section méthodologie de ce rapport.

Constatation 1 : Dans toutes les évaluations confondues, la majorité des critères d'appréciation de la qualité ont été notés « Satisfaisant » et « Très satisfaisant ».

Un peu plus de 70 % des critères d'appréciation de la qualité ont reçu la note « Satisfaisant » ou « Très satisfaisant » avec une répartition assez homogène de ces notes parmi les quatre groupes de critères, montrant seulement une différence de 12,6 % entre les chiffres les plus bas et les chiffres les plus élevés (voir le graphique 4.1). Les critères liés au contexte, à l'objet, à la finalité, aux objectifs et à la portée de l'évaluation ont reçu les meilleures notes (75,3 % de notes « Satisfaisant » ou « Très satisfaisant ») par rapport aux autres groupes, suivis de très près par les critères de clarté et de fiabilité des rapports (75 %). En moyenne, les critères liés aux processus ont obtenu moins de bonnes notes.

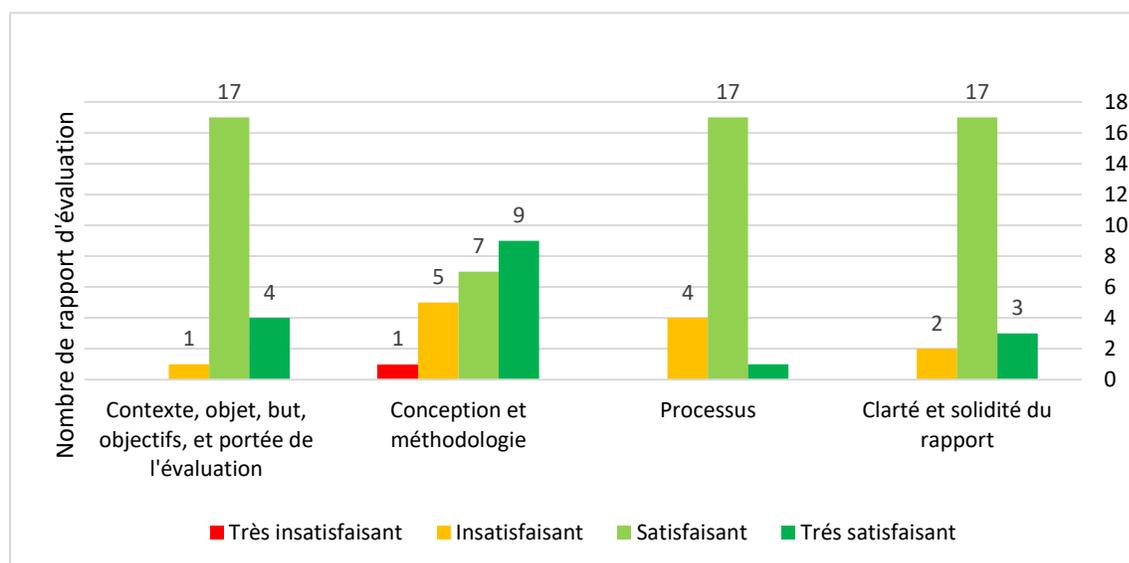
Graphique 4.1 Pourcentage des critères notés 3 et 4 par groupe, toutes évaluations confondues



Constatation 2 : La note la plus fréquente pour les quatre groupes de critères était « Satisfaisant », et c'est le groupe de la conception et de la méthodologie qui montrait la plus grande variation de notes.

La note la plus fréquente enregistrée était « Satisfaisant », traduisant une image globalement positive de la qualité des évaluations examinées (voir le graphique 4.2). Les groupes 1 et 4 ont reçu les notes les plus élevées, avec 17 évaluations jugées satisfaisantes en moyenne pour ces groupes, alors que le groupe 2 montre la plus grande variation en termes de notes moyennes.

Graphique 4.2 Notes moyennes des groupes, toutes évaluations confondues (n=22)



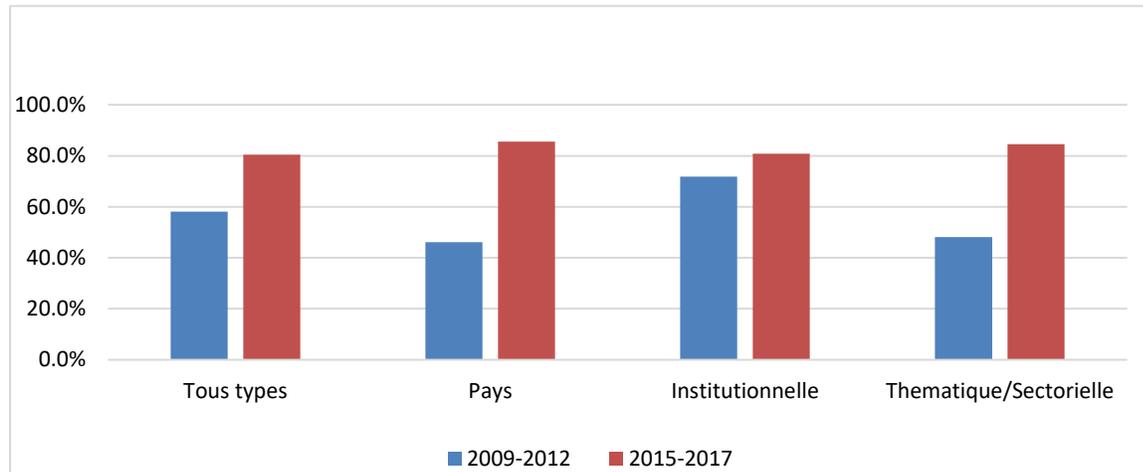
4.1.2 Comparaison globale entre les types d'évaluations entre la première et la deuxième période

Cette section compare les critères et les notes moyennes des groupes entre les première et deuxième périodes et entre les types d'évaluations.

Constatation 3 : On observe une nette amélioration de la qualité des évaluations au fil du temps de 2009-2012 à 2015-2017.

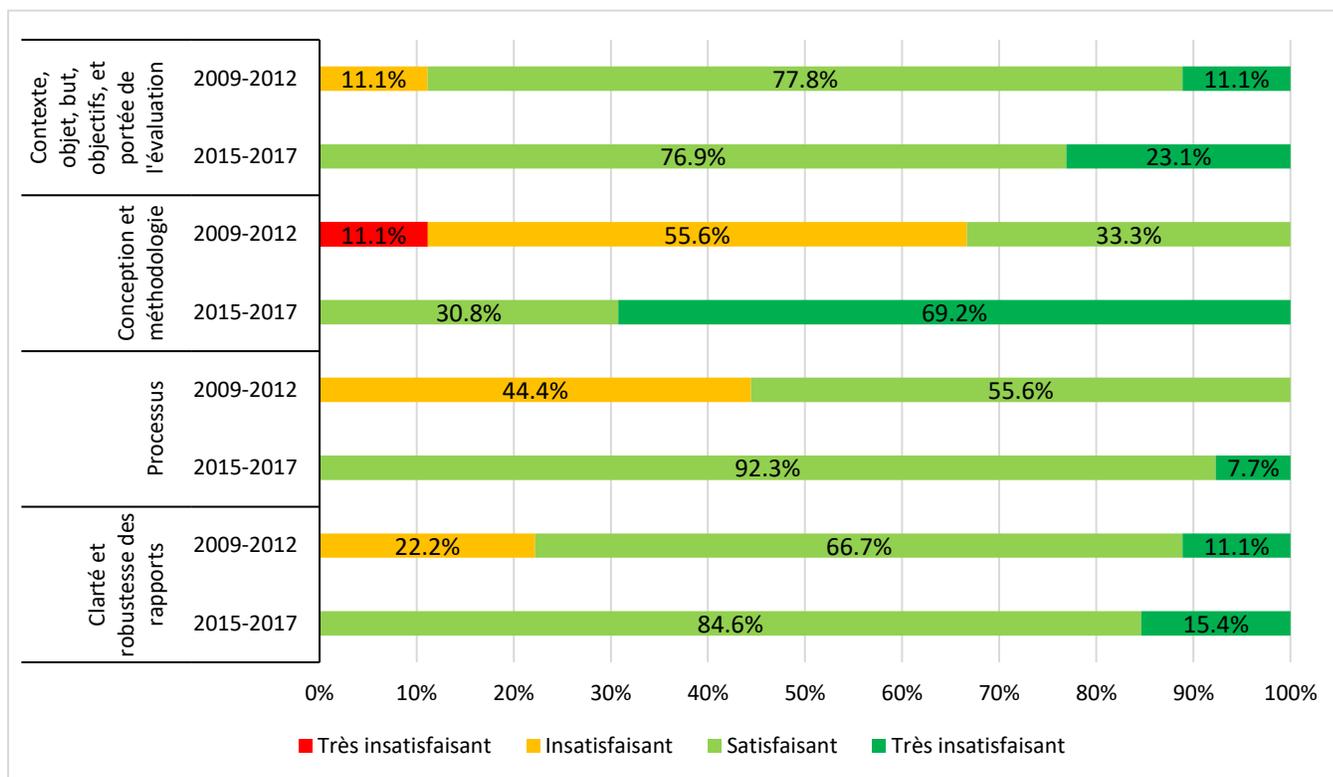
Si on compare les appréciations des évaluations entre la première et la deuxième période, la qualité s'améliore nettement avec le temps, les évaluations de pays et thématiques/sectorielles montrant une amélioration plus nette de la qualité que les évaluations institutionnelles, simplement parce que ces évaluations ont été mieux notées que les autres types d'évaluations pour la première période (voir le graphique 4.3). Il convient de préciser que la première évaluation de synthèse de la première période n'a pas pu être comparée par période aux quatre autres évaluations examinées dans la catégorie « Autres ».

Graphique 4.3 Comparaison du pourcentage des critères notés 3 et 4 par type et période d'évaluation



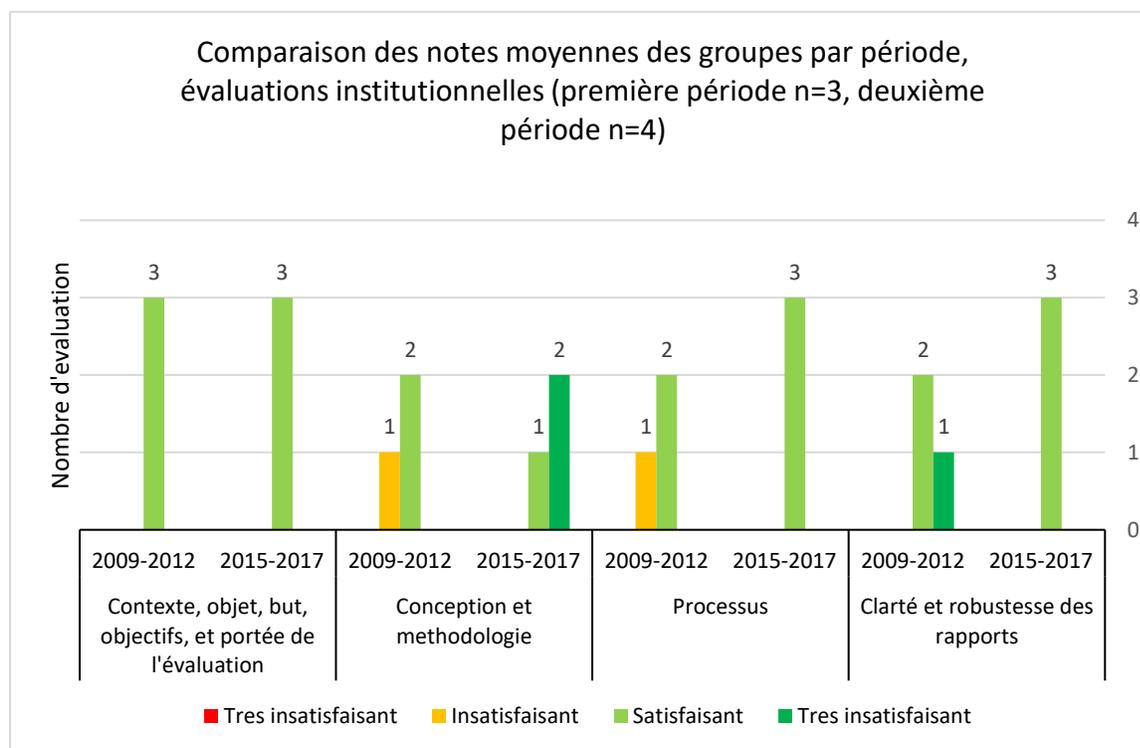
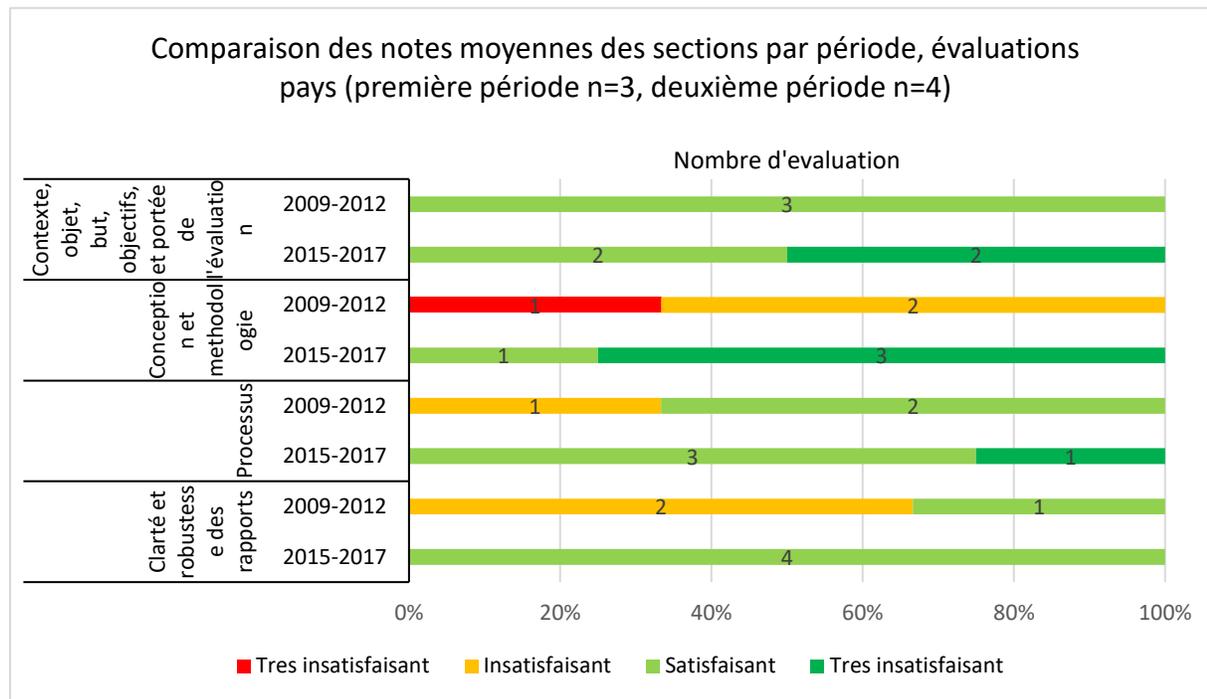
Alors que 11,1% à 55,6% des évaluations de la première période ont reçu une note moyenne « Insatisfaisant » pour chaque groupe de critères, les 13 évaluations de la deuxième période ont obtenu une note moyenne « Satisfaisant » ou « Très satisfaisant » pour tous les groupes (voir graphique 4.4). Le pourcentage de notes « Satisfaisant » et « Très satisfaisant » pour chacun des 26 critères dans toutes les évaluations est présenté à l'annexe III.

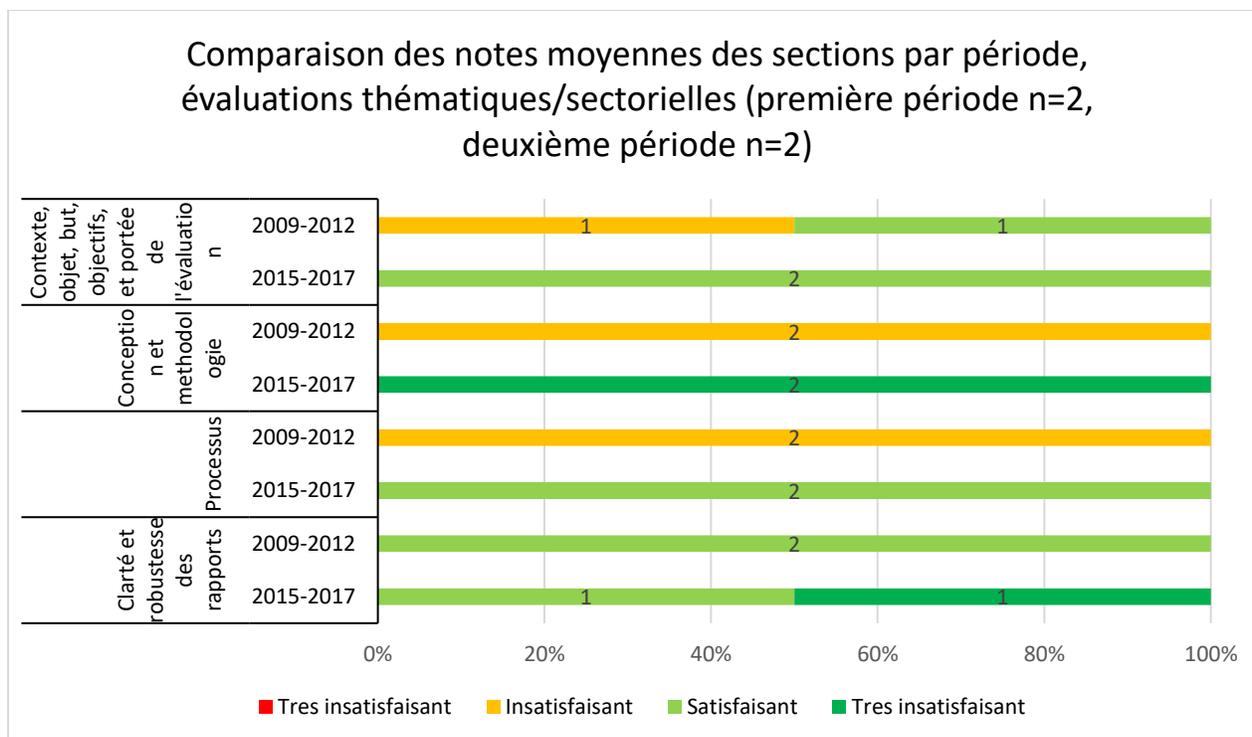
Graphique 4.4 Comparaison des notes moyennes des groupes entre la première et la deuxième période, toutes évaluations confondues (première période n=9, deuxième période n=13)



Avec les différents types d'évaluation (à l'exception de celles de la catégorie « Autres ») représentés de manière presque égale en nombre dans les deux périodes, on constate une amélioration notable de tous les types d'évaluation, de la première à la deuxième période (voir graphique 4.5).

Graphique 4.5 Comparaison des notes moyennes des groupes par type d'évaluation, première et deuxième période





4.2 Analyse comparative des forces et des faiblesses des évaluations selon les critères entre la première et la deuxième période

Cette section compare les forces et les faiblesses des évaluations examinées pour chacune des deux périodes, faisant ressortir les thèmes émergents à travers l'appréciation plus spécifique des critères individuels dans chacun des quatre groupes.

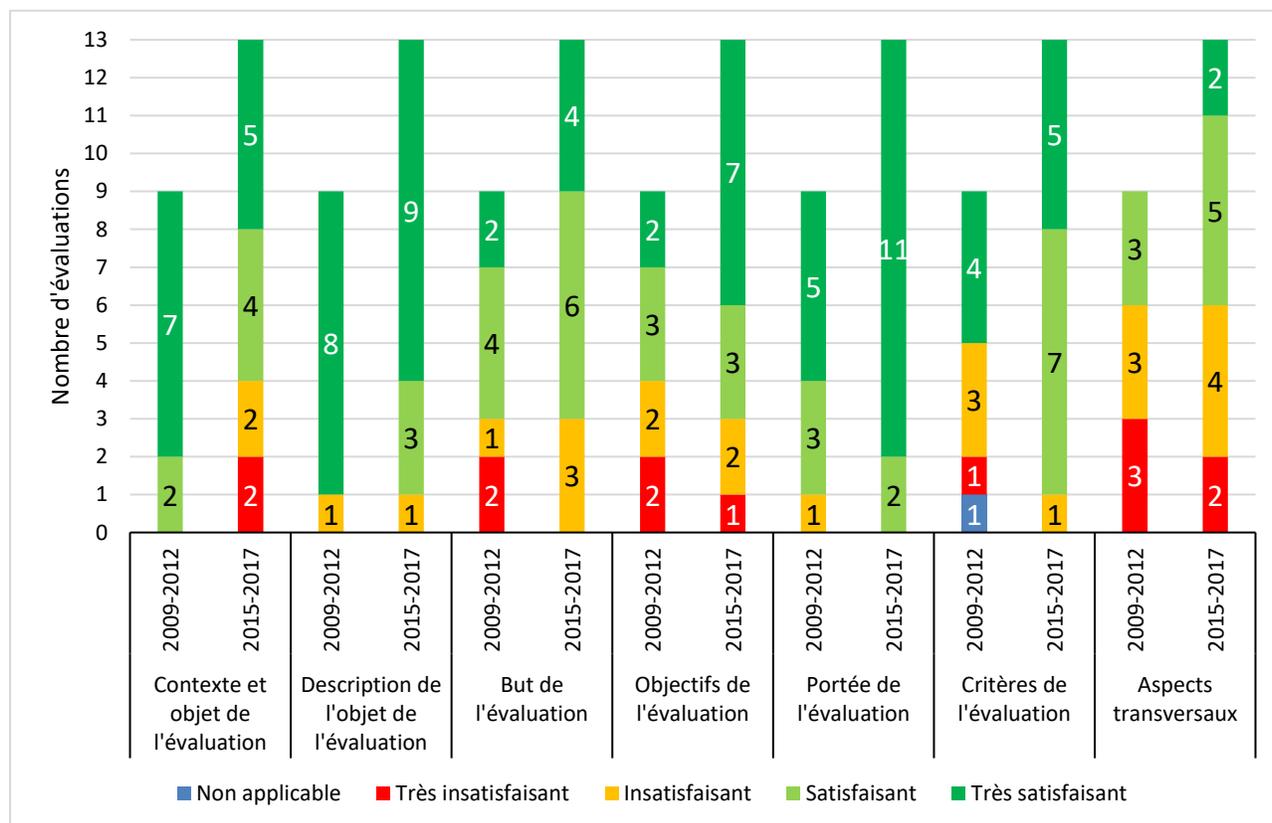
4.2.1 Contexte, objet, finalité, objectifs et portée des évaluations

Constatation 4 : La description de l'objet de l'évaluation et la portée de l'évaluation sont les critères les mieux notés, et les questions transversales représentent le critère le moins bien noté, au sein de ce groupe au cours des deux périodes considérées.

Parmi les sept critères liés au contexte, à l'objet, à la finalité, aux objectifs et à la portée de l'évaluation, la description de l'objet et de la portée de l'évaluation ont été les critères les mieux notés en moyenne (voir le graphique 4.6). En comparant les notations des critères individuels entre les évaluations de première et de la deuxième période, on observe que les évaluations de la première période présentent des notations légèrement plus élevées pour ce qui est de la description du contexte de l'objet de l'évaluation, les évaluations de la deuxième période montrent, elles, une amélioration notable des notes relatives à l'identification claire de critères d'évaluation en conformité avec la finalité de l'évaluation.

La plupart des évaluations ont porté sur le calendrier de l'objet de l'évaluation, ses principaux objectifs, ses stratégies et son budget. Parmi les bonnes pratiques identifiées, on retient la manière dont certaines évaluations comprenaient une description détaillée de l'objet de l'évaluation et de son contexte. Par exemple, certaines évaluations ont fourni des détails sur des évaluations similaires effectuées auparavant et sur la manière dont la Banque avait réagi à ces appréciations.

Graphique 4.6 Notations du groupe 1 : comparaison entre périodes (première période n=9, deuxième période n=13)



Les questions transversales dans ce groupe ont reçu la note la plus faible en moyenne, avec près de 50 % (6 sur 13) des évaluations de la deuxième période jugées insatisfaisantes ou très insatisfaisantes et seulement 2 évaluations de seconde période sur 13 (15 %) ont été jugées très satisfaisantes, selon ce critère. La principale raison pour laquelle ce critère a été noté « Insatisfaisant », c'est que les questions transversales n'étaient pas abordées dans la description de l'objet de l'évaluation et de son contexte, ni dans la portée de l'évaluation. Par exemple, en ce qui concerne l'égalité des sexes, certaines évaluations ne traitaient que partiellement cette question en incluant des indicateurs ventilés par sexe dans la matrice d'évaluation, mais n'incluaient pas d'explication narrative sur la manière dont la question de l'égalité des sexes serait abordée dans le processus d'évaluation. De bons exemples ont été remarqués, comme les évaluations fournissant des informations contextuelles sur des questions transversales dans la description de l'historique et de l'objet de l'évaluation, celles prenant en considération les questions transversales dans les questions d'évaluation, ou encore celles intégrant une déclaration d'intention explicite afin de traiter les questions transversales dans le processus d'évaluation.

La majorité des évaluations ont été notées « Satisfaisant » et « Très satisfaisant » en ce qui concerne la description de la finalité et des objectifs de l'évaluation, avec une amélioration notable observée à l'égard de ces critères au fil du temps de la première à la deuxième période. Cependant, il convient de mentionner que plusieurs évaluations, même dans la deuxième période, n'apportaient aucune précision sur la pertinence de leur calendrier, les utilisateurs visés ou l'utilisation prévue dans la présentation de leur finalité. De plus, certaines évaluations ne distinguaient pas clairement la finalité et les objectifs de l'évaluation.

En termes de variance par type d'évaluation, les évaluations par pays sont légèrement mieux notées que les autres types d'évaluation. C'est d'ailleurs le seul type d'évaluation ayant des notes « Très satisfaisantes » pour ce groupe. Les évaluations thématiques/sectorielles sont légèrement moins bien notées, car elles incluent une évaluation jugée insatisfaisante dans ce groupe (voir le graphique 4.5 à la section 4.1.2).

4.2.2 Conception et méthodologie

Constatation 5 : Au sein de ce groupe, la collecte de données, l'analyse et l'échantillonnage constituent le critère le mieux noté, puisqu'aucune évaluation n'a reçu une note « Très insatisfaisant » sur ce critère, et l'alignement des questions d'évaluation avec les critères et les indicateurs est le critère le moins bien noté, bien qu'on enregistre une amélioration significative de ce critère dans les évaluations de la deuxième période.

En ce qui concerne les critères de conception et de méthodologie, le plus grand nombre de notes « Satisfaisant » et « Très satisfaisant » se rapportaient à la collecte de données, à l'analyse et à l'échantillonnage (voir le graphique 4.7). Cependant, près de la moitié des évaluations de pays, toutes issues de la première période, présentaient une description des méthodologies assez médiocre en général (voir le graphique 4.5, section 4.1.2).

Les problèmes relevés étaient, entre autres, une mauvaise identification de la conception de l'évaluation, l'absence d'une matrice d'évaluation ou d'une théorie du changement (TdC), une description médiocre des méthodes d'échantillonnage ou de l'échelle de notation utilisée pour apprécier la qualité.

Exemple de cas : Politique de gestion des ressources humaines et orientations stratégiques de la Banque africaine de développement : Une évaluation formative

Le modèle de conception de l'évaluation formative et théorique est clairement décrit et aligné de manière appropriée avec la finalité. Une stratégie détaillée de collecte et d'analyse de données mixtes est décrite avec un échantillon complet de sources de données. Les questions d'évaluation pertinentes sont décrites dans une matrice d'évaluation détaillée en annexe. Les limites de la méthode et les stratégies d'atténuation sont décrites de manière appropriée.

Les améliorations les plus notables au fil du temps parmi les critères de ce groupe concernent l'alignement des questions d'évaluation avec les critères et les indicateurs d'évaluation dans une matrice d'évaluation et l'identification des risques et/ou des limites d'évaluation (voir le graphique 4.7). Toutes les évaluations de la première période et 23 % des évaluations (3 sur 13) de la deuxième période n'ont pas présenté de manière satisfaisante une matrice d'évaluation. Dans certains cas récents, où une matrice d'évaluation a été incluse, certains éléments clés n'y apparaissaient pas, notamment les questions en lien avec les indicateurs appropriés ou les méthodes et les sources de collecte de données.

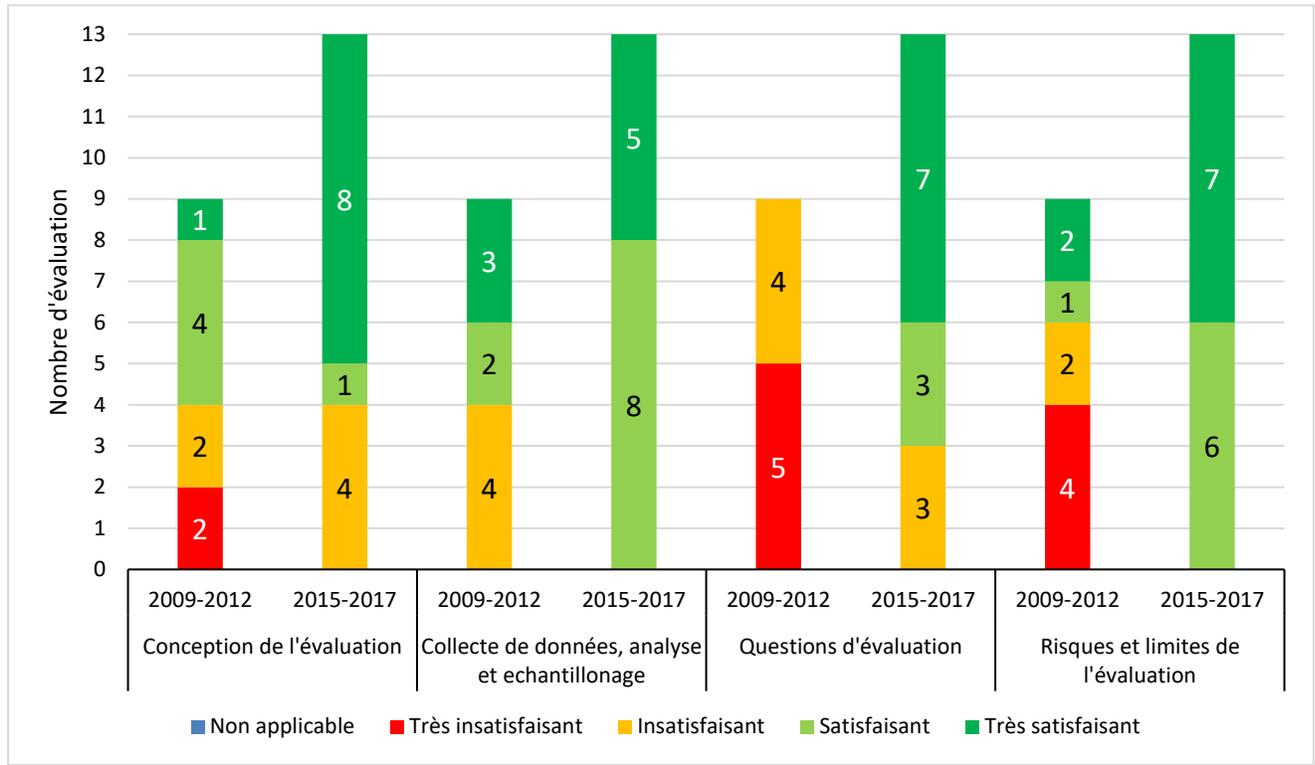
Cela concernait en particulier les évaluations de la catégorie « Autres ». Bien que toutes les évaluations de la deuxième période aient été jugées satisfaisantes pour ce qui est de la collecte, de l'échantillonnage et de l'analyse des données, ainsi que de l'identification des limites d'évaluation et des stratégies d'atténuation, environ 30 % (4 sur 13) n'ont pas établi de manière satisfaisante le modèle de conception de l'évaluation.

L'échelle de notation utilisée par les évaluations d'IDEV a également été jugée problématique. En se référant aux exemples d'une évaluation de pays et d'une évaluation institutionnelle, il semble que beaucoup d'efforts aient été déployés pour expliquer comment les notations pour chaque niveau de l'échelle de notation à six points étaient définies (par exemple, évaluation EGRD, annexes D, G, et H ; évaluation du Ghana, annexes C et E).¹¹ Sans ces définitions cependant, les notations seraient moins claires puisque l'on pourrait techniquement considérer les notes intermédiaires (MI et MS) comme équivalents. Cela dit, même lorsque ces notes étaient définies, elles manquaient parfois de clarté en ne définissant pas clairement les limites supérieures des fourchettes.¹² Ces définitions plus fines, qui spécifient une fourchette quantitative, peuvent être plus pertinentes pour évaluer des facteurs ou des critères essentiellement quantitatifs, tels que l'efficacité. Il convient de noter qu'IDEV a récemment remplacé l'échelle de notation à 6 points par une échelle d'évaluation à 4 points avec les intitulés suivants : « Très insatisfaisant », « Insatisfaisant », « Satisfaisant », « Très satisfaisant ». Cependant, bien que moins lourde que l'échelle précédente à six points, cette échelle est également problématique puisqu'elle ne permet pas une appréciation nuancée entre « Insatisfaisant » et « Satisfaisant ».

¹¹ Échelle de notation à six points : très insatisfaisant (TI), insatisfaisant (I), modérément insatisfaisant (MI), modérément satisfaisant (MS), satisfaisant (S), très satisfaisant (TS).

¹² Par exemple, les définitions des notes MI et MS en lien avec l'évaluation du critère de pertinence dans l'annexe D de l'évaluation de synthèse de l'EGRD.

Graphique 4.7 Notations du groupe 2 : comparaison entre périodes (première période n=9, deuxième période n=13)



4.2.3 Processus

Constatation 6 : Les liens entre les produits d'évaluation pour chaque évaluation étaient logiques dans l'ensemble, tandis que les principes d'évaluation (indépendance, impartialité, éthique, coordination et alignement, partenariat et renforcement des capacités) n'étaient pas présentés de manière satisfaisante pour les deux périodes.

Alors que les notes de qualité liées au processus étaient assez variables (voir le graphique 4.8), les liens entre les produits d'évaluation ont été jugés globalement positifs, 89 % (8 sur 9) et 85 % (11 sur 13) des évaluations de la première et de la deuxième période ayant été notées « Satisfaisantes » ou mieux encore sur ce critère. Certaines répétitions d'informations ont été observées dans les documents d'orientation, les rapports de démarrage et parfois dans les termes de référence pour les mêmes évaluations, mais le personnel d'IDEV a expliqué que tous ces produits n'étaient pas nécessaires à toutes les évaluations ; ils ont chacun un objectif distinct en fonction du contexte de l'évaluation et de la mesure dans laquelle les consultants externes sont impliqués. Toutefois, IDEV ne dispose pas de directives spécifiques sur ce qui doit apparaître a minima dans chaque type de document.

Certaines évaluations ont été très bien notées lors de l'examen des principes d'évaluation, en particulier le partage des connaissances par la diffusion des résultats. Cependant, alors que la politique d'IDEV met l'accent sur les principes directeurs d'indépendance, de crédibilité, d'impartialité, de transparence et de partenariat¹³, ces principes, en plus de la coordination et de l'alignement, et du renforcement des capacités, n'étaient pas bien ou pas explicitement pris en compte dans l'ensemble des évaluations examinées. Ces principes peuvent avoir été appliqués dans la pratique, mais aucune preuve tangible ne vient en attester. Le nombre d'évaluations jugées satisfaisantes ou plus pour ce qui est de l'intégration des autres principes d'évaluation de l'OCDE-CAD et d'IDEV a augmenté de 15 % (2 sur 13) pour la deuxième période, contre zéro pour la première période, mais la plupart des évaluations sont toujours jugées comme insatisfaisantes ou pire dans cette catégorie. Très peu d'évaluations se réfèrent à de tels principes d'évaluation et, lorsqu'elles le font, la référence est seulement générique et il est impossible de voir clairement comment les principes ont été appliqués concrètement.

En ce qui concerne l'assurance qualité, 78 % (7 sur 9) des évaluations effectuées au cours de la première période ont été jugées insatisfaisantes sur ce critère, simplement parce que les mécanismes d'assurance-qualité n'étaient pas abordés dans l'ensemble des documents examinés. À l'inverse, 92 % (12 sur 13) des évaluations menées au cours de la deuxième période ont été jugées satisfaisantes ou mieux encore sur la base de ce critère. Toutes les évaluations (sauf une) ont décrit le processus d'assurance-qualité. Les évaluations qui ont reçu la note « Très satisfaisant » avaient des mécanismes internes et externes d'assurance-qualité, comme des comités d'examen par les pairs ou des experts indépendants qui ont examiné et commenté les produits livrables de l'évaluation.

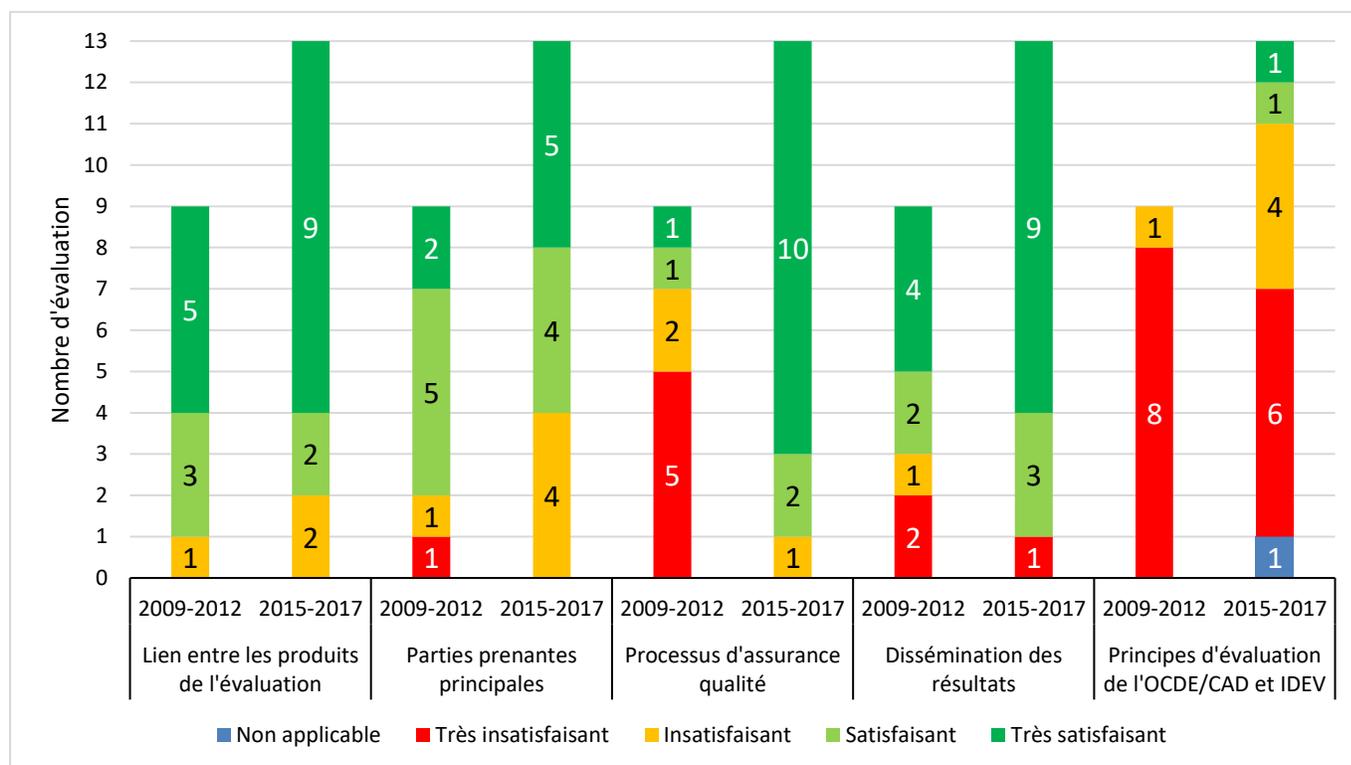
Une tendance opposée a été observée pour l'engagement des principales parties prenantes entre les première et deuxième périodes, avec 78 % (7 sur 9) des évaluations de la première période et 69 % (9 sur 13) des évaluations de la deuxième période jugées satisfaisantes ou meilleures encore. Les évaluations notées « Très satisfaisant » sur ce critère ont décrit le processus d'engagement des parties prenantes dans la conception de l'évaluation, ainsi que dans la collecte des données et la validation des constatations. Étant donné l'importance de la participation des parties prenantes à l'évaluation, il est difficile de savoir pourquoi les évaluations les plus récentes n'ont pas toujours systématiquement identifié les parties prenantes ni leurs rôles et contributions dans le processus.

Exemple de cas : Assurance-qualité

L'Évaluation Globale des Résultats de Développement du groupe de la Banque africaine de développement 2004-2013, publiée en 2016, a été identifiée comme recelant de bonnes pratiques car elle décrit l'assurance-qualité interne et externe. En interne, la revue des facteurs contextuels (RFC) et l'appréciation des résultats de projet ont été examinées par deux membres du personnel et coordonnées au niveau central pour assurer la cohérence. En externe, un groupe de conseillers indépendants de haut niveau a été appelé à fournir des conseils méthodologiques pendant la phase de conception, ainsi que des commentaires sur le projet et le rapport final. Ce groupe a également fourni une lettre, incluse dans le rapport d'évaluation, qui atteste de la qualité de l'évaluation, confirmant que les évaluateurs avaient répondu à leurs commentaires.

¹³ Politique d'évaluation indépendante, Groupe de la Banque africaine de développement, juillet 2016, pp. 2-4.

Graphique 4.8 Notations du groupe 3 : comparaison entre périodes (première période n=9, deuxième période n=13)



Enfin, en ce qui concerne la diffusion des résultats, 92 % (12 sur 13) des évaluations menées au cours de la deuxième période ont obtenu une note satisfaisante ou meilleure encore, contre 67 % (6 sur 9) pour la première période. Dans les évaluations récentes, l'inclusion d'un plan de diffusion identifiant les publics cibles pour divers produits de communication tels que des mémoires, des rapports de synthèse ou des présentations PowerPoint est considérée comme une bonne pratique. Certains plans de diffusion comprennent également un calendrier et une personne responsable de la conduite des activités d'évaluation.

4.2.4 Clarté et fiabilité des rapports

Constatation 7 : La grande majorité des rapports d'évaluation étaient bien structurés et comprenaient un résumé analytique concis, présentaient des constatations fondées sur des données probantes, respectaient les critères et les questions d'évaluation et introduisaient des recommandations pertinentes. Cependant, les sources desquelles les données ont été extraites n'étaient souvent pas explicitement identifiées dans les évaluations des deux périodes.

Environ 75 % des rapports d'évaluation de la première période et tous les rapports d'évaluation de la deuxième période ont été jugés satisfaisants selon les critères relatifs à la structure du rapport, au résumé analytique, à la présentation des constatations fondées sur des données probantes qui traitaient des critères et questions d'évaluation, et à la clarté des recommandations (voir le graphique 4.9).

Cependant, dans les résultats du rapport, dans tous les types d'évaluations et dans les deux périodes, bien que des sources documentaires exhaustives aient été étudiées et que certains intervenants clés aient été consultés dans le processus d'évaluation, les groupes d'intervenants à partir desquels les données probantes étaient dérivées n'étaient souvent pas identifiés, de telle sorte qu'il était souvent difficile de savoir quelles étaient leurs opinions communes ou distinctes et dans quelle mesure les méthodes et les sources de collecte de données étaient triangulées.

La plupart des rapports traitaient systématiquement des critères et des questions d'évaluation et incluaient des recommandations pertinentes. Cependant, les recommandations ne précisaient généralement pas un calendrier précis de mise en œuvre indiquant le niveau de priorité de chaque recommandation.

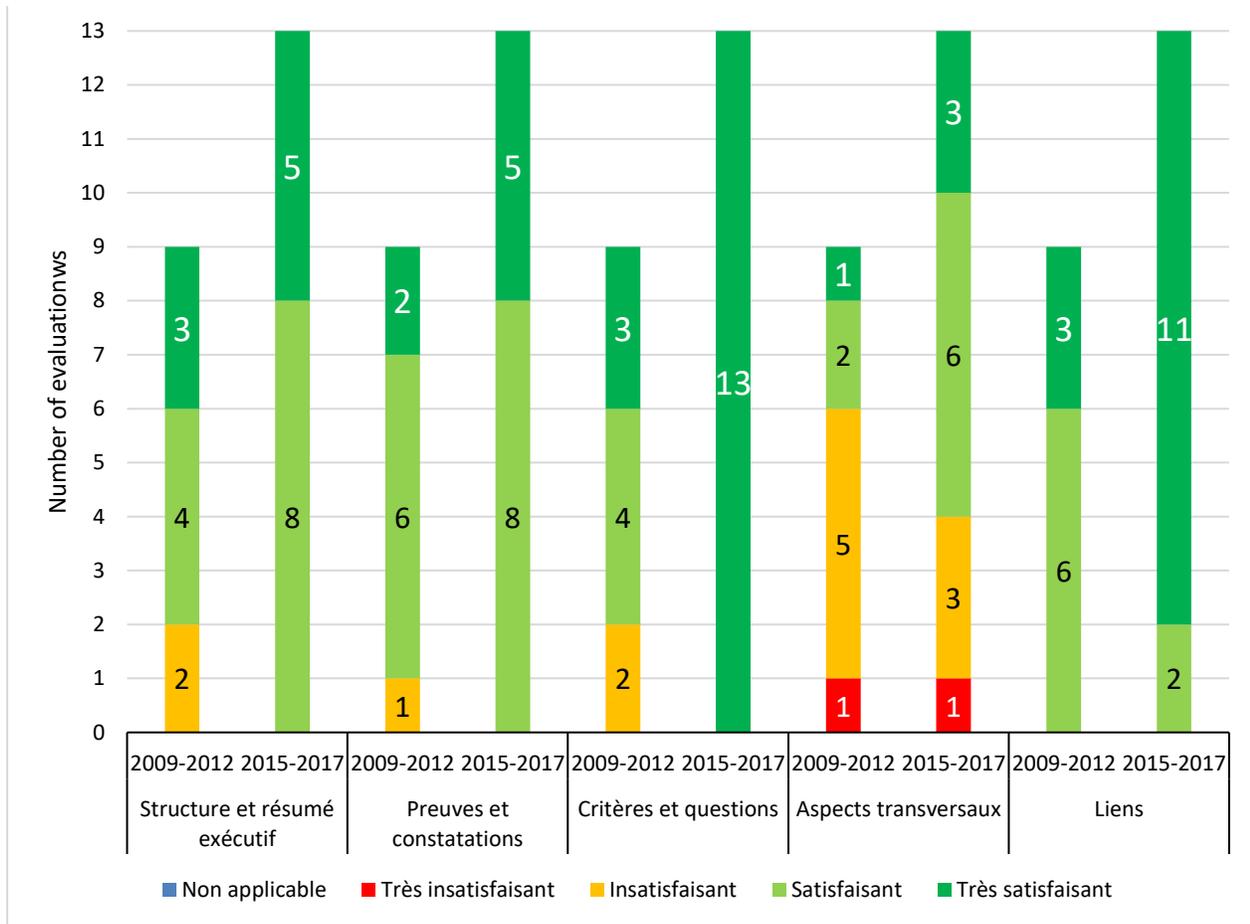
Presque toutes les évaluations de la deuxième période comprennent une réponse de la direction¹⁴ et toutes ces réponses examinées indiquent un accord global avec les recommandations faites.

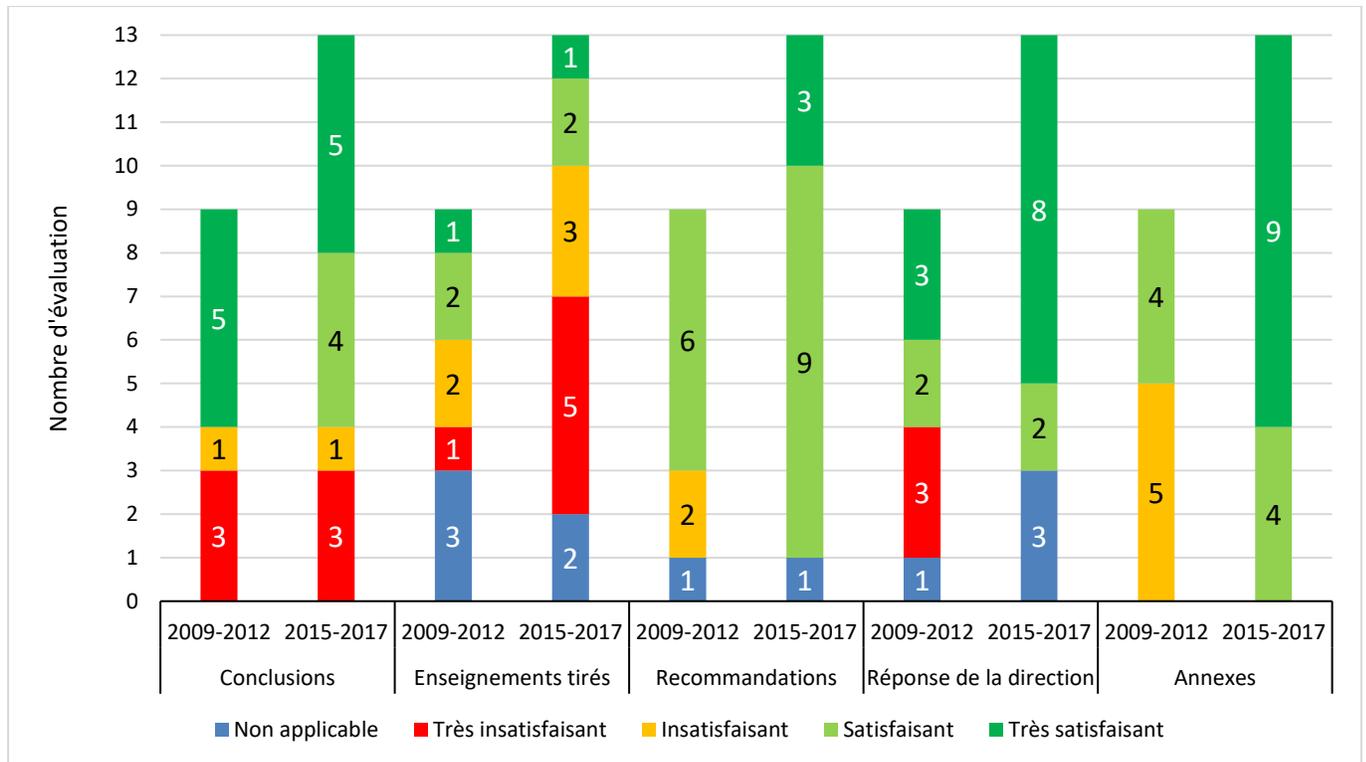
Exemple de cas : Afrique du Sud - Évaluation de la stratégie et du programme pays 2004–2015

La preuve repose sur de multiples sources de données, y compris un examen du portefeuille et des consultations avec les intervenants clés. Dans le rapport de synthèse, les sources de données sont systématiquement présentées dans l'ensemble du rapport et les points de vue des différentes parties prenantes permettent de dresser un tableau nuancé de la performance de la Banque. La manière dont l'information a été triangulée est claire. Les lacunes dans les données sont également abordées. Par exemple, le rapport de synthèse confirme qu'il n'y avait pas suffisamment d'informations pour démontrer si la Banque se gérait de manière rentable.

Graphique 4.9 Notations du groupe 4 : comparaison entre périodes (première période n=9, deuxième période n=13)

¹⁴ Même s'il est possible qu'il existe une réponse de la direction pour toutes les évaluations examinées, certaines sont restées introuvables. Ces évaluations n'ont donc pas pu recevoir une note satisfaisante sur ce critère.





Constatation 8 : Toutes les évaluations ont démontré des liens clairs entre les données probantes, les constatations, les conclusions, les leçons apprises et les recommandations générales, mais les constatations et les conclusions ne reflétaient pas bien les questions transversales dans presque la moitié des rapports examinés dans les deux périodes. En outre, les leçons apprises, le cas échéant, n'étaient pas incluses ou pas correctement identifiées dans la moitié des rapports, en particulier ceux de la deuxième période.

Il existe des liens évidents entre les constatations, les conclusions et les recommandations générales, même si dans certains cas, les conclusions et recommandations n'abordent pas toute la gamme des principales constatations ou conclusions résumant simplement certaines constatations clés sans y ajouter de valeur significative (voir le graphique 4.9). Les faiblesses les plus observées dans ce groupe de critères et à travers les deux périodes concernent l'appréciation des questions transversales et la formulation des leçons apprises le cas échéant.

Tandis que certaines questions, telles que l'égalité des sexes, la fragilité, l'environnement et la croissance inclusive sont clairement considérées comme importantes dans la stratégie décennale de la Banque pour 2013-2022¹⁵, il semble qu'il y ait eu une occasion manquée d'aborder ces questions de manière transversale dans les évaluations examinées, y compris dans les évaluations de la deuxième période qui ont été menées après la publication de la stratégie décennale. Par exemple, bien que les constatations sur les questions transversales aient pu être présentées dans les rapports d'évaluation, l'appréciation de celles-ci n'a pas toujours été reprise dans les conclusions, les leçons apprises et les recommandations, le cas échéant. Dans certains cas, quelques questions transversales ont été identifiées de diverses manières

¹⁵ BAD, *Au centre de la transformation de l'Afrique : Stratégie pour 2013-2022*. 2013, pp. 1-3.

dans les rapports, de sorte qu'il n'était pas toujours très évident d'identifier avec précision les problèmes intersectoriels. En outre, dans certaines évaluations où il aurait été pertinent de traiter de certains thèmes transversaux, ces derniers ont été inexplicablement ignorés. L'évaluation d'impact est un exemple pertinent de ce point de vue. Alors que la croissance économique inclusive et l'égalité des sexes ont été évaluées en termes d'impact relatif aux interventions eau/EAH sur l'emploi des femmes en Éthiopie, et mesurées par une enquête communautaire, d'autres questions transversales, telles que celles liées à l'environnement, n'ont pas encore été abordées et sont clairement pertinentes pour les interventions dans le secteur de l'eau.

Exemple de cas : Appui de la Banque africaine de développement au développement des chaînes de valeur agricoles : Enseignements pour la stratégie « Nourrir l'Afrique »

Les rapports techniques et de synthèse sont bien écrits, concis et clairement structurés. Ils comprennent un résumé analytique qui contient toutes les informations pertinentes tirées de l'évaluation. Les données présentées semblent provenir de sources multiples, bien que les groupes de parties prenantes (sources primaires uniquement) ne soient pas toujours clairs. Les critères d'évaluation sont respectés et les questions trouvent systématiquement une réponse. Les constatations traitent en profondeur des questions transversales liées à l'égalité des sexes, à l'inclusivité et à la durabilité environnementale, bien que les questions environnementales puissent être traitées avec plus d'attention dans les recommandations. Les conclusions sont déduites des constatations et fournissent des indications supplémentaires. Les leçons apprises de l'évaluation se font toutefois rares. Les recommandations sont pertinentes.

En ce qui concerne les leçons apprises, plusieurs évaluations de pays et institutionnelles en particulier ne les ont pas traitées de manière adéquate. Dans les cas où les évaluations ne les incluaient pas et ne visaient pas spécifiquement à en fournir, c'est-à-dire lorsque les leçons n'étaient pas spécifiées dans l'énoncé de la finalité de l'évaluation, ce critère était jugé non applicable, ce qui n'affectait pas la notation de ce groupe. Cependant, dans les cas où l'évaluation identifiait des leçons dans l'énoncé de ses objectifs, mais qu'aucune d'entre elles n'était identifiée dans le rapport de synthèse, la note attribuée à ce critère en était affectée négativement. Les notes ont également été affectées négativement lorsque les leçons apparaissaient certes dans le rapport, mais qu'elles n'étaient pas clairement ou correctement formulées pour optimiser leur pertinence.

5 Conclusions, leçons apprises et recommandations

5.1 Conclusions

Sur la base des résultats de cette appréciation de la qualité, la qualité des évaluations s'est améliorée depuis la mise en œuvre de la stratégie d'évaluation d'IDEV. Cependant, il est important de noter que nous ne pouvons pas attribuer cette amélioration à la stratégie elle-même, étant donné les multiples facteurs qui peuvent influencer la qualité. Alors que les notes des évaluations par pays et des évaluations

thématiques/sectorielles ont montré la plus forte progression (vers les notes « Satisfaisant » et « Très satisfaisant ») entre les deux périodes (augmentation de près de 40 % dans le cas des évaluations par pays), la qualité des évaluations institutionnelles, déjà considérablement plus élevée que celle des évaluations thématiques/sectorielles pour la période 2009-2012, a augmenté dans une moindre mesure au cours de la période 2015-2017.

La majorité des critères d'appréciation de la qualité ont été jugés satisfaisants et très satisfaisants, en particulier dans le cas des évaluations effectuées au cours de la deuxième période. Cependant, il est encore possible d'améliorer les résultats dans certains domaines. Des faiblesses ont notamment été observées dans les détails fournis sur la conception de l'évaluation et sur la présence d'une matrice d'évaluation complète avec toute la méthodologie, décrivant les parties prenantes et leur implication dans le processus d'évaluation, abordant explicitement l'ensemble des principes d'évaluation, intégrant efficacement les questions transversales et formulant clairement les leçons apprises.

Les questions transversales, telles que l'égalité des sexes, la croissance inclusive, l'environnement/le changement climatique et la fragilité, qui sont explicitement mentionnées dans la stratégie décennale de la BAD ne sont pas toujours clairement identifiées dans les évaluations examinées ; et quand elles le sont, leur analyse manque de profondeur. Les constatations révèlent une couverture inégale de ces dimensions, à la fois pour chaque type de produit d'évaluation et pour l'ensemble des produits. Ces dimensions sont communément appelées « transversales », mais elles ne sont pas toujours efficacement intégrées de manière transversale dans les évaluations.

De même, la couverture des principes d'évaluation reste inégale parmi les évaluations réalisées au cours de la deuxième période. Alors que les critères d'assurance-qualité et de partage des connaissances par la diffusion des résultats ont été notés, les principes d'impartialité, d'éthique, de coordination et d'alignement, de partenariat et de renforcement des capacités n'ont pas été explicitement abordés dans les produits d'évaluation examinés. L'engagement des principales parties prenantes n'est souvent pas clairement décrit. Bien que ces principes puissent être bel et bien considérés dans la pratique, ils ne sont pas suffisamment documentés.

En ce qui concerne la méthodologie d'évaluation, les chiffres présentés indiquent que la conception d'évaluation n'est pas toujours clairement décrite et que la matrice d'évaluation n'est pas présentée de manière adéquate.

Les constatations rapportées semblent être fondées sur des données probantes et appuyées par des données secondaires exhaustives, mais des sources primaires limitées. Certains chevauchements ont été notés entre les informations dans les différentes sections des rapports, par exemple, les constatations incluant les jugements des évaluateurs et les notes de performance, les conclusions reprenant essentiellement certaines constatations clés et les leçons apprises formulées comme des conclusions, voire même des recommandations. De même, pour ce qui relève des processus, un certain chevauchement a également été observé entre certains produits d'évaluation, par exemple, les documents d'orientation et les termes de référence.

L'échelle de notation à 6 points, habituellement utilisée par IDEV pour les évaluations de l'échantillon, était astreignante et trop spécifique, conduisant à une micro-analyse qui n'est ni pratique ni particulièrement pertinente, étant donné la portée assez large des évaluations menées par IDEV, en termes de couverture géographique et de délai. Alors que cette échelle a récemment été remplacée par une échelle à 4 points, cette nouvelle échelle est également problématique dans une certaine mesure car les notes utilisées ne permettent pas de donner une note entre « Insatisfaisant » et « Satisfaisant ».

5.2 Leçons apprises

Les leçons suivantes ont été apprises de cette appréciation de la qualité :

Leçon 1 : Le test des outils de collecte de données est une étape nécessaire du processus d'appréciation/d'évaluation, à la fois pour s'assurer que les outils sont adaptés à la situation et que tous les collecteurs de données utilisent les outils de manière cohérente. Cela facilite la compilation et l'analyse ultérieures des données et renforce la validité des résultats.

Leçon 2 : Une approche d'équipe plutôt qu'une appréciation individuelle unique des évaluations renforce la cohérence dans les notations et les jugements.

Leçon 3 : L'implication des principales parties prenantes dans la discussion et/ou l'élaboration des recommandations renforce leur appropriation et la probabilité que ces dernières soient effectivement mises en œuvre.

Leçon 4 : Imposer des limites strictes au nombre maximum de pages pour les rapports d'évaluation (dans ce cas, les rapports de synthèse), en particulier dans les évaluations complexes, telles que celles produites par IDEV, peut empêcher l'ajout de certains éléments importants nécessaires à une bonne compréhension de l'évaluation. Cela peut empêcher le lecteur de bien comprendre le processus et les problèmes d'évaluation et affecter la qualité des rapports.

Leçon 5 : Disposer de directives claires sur les éléments de base qui devraient être inclus dans chaque produit d'évaluation (et chacune des sections principales des différents types de produits) permet de clarifier les exigences en termes de rédaction des rapports et contribue à assurer une qualité plus uniforme des documents produits.

5.3 Recommandations

Sur la base des constatations et des conclusions de cette appréciation de la qualité, nous présentons quelques recommandations relatives aux différents produits d'évaluation pour la BAD-IDEV dans leur soutien de la fonction d'évaluation.

Recommandation 1 : Considérer l'ajout d'une cartographie des parties prenantes. Cette exigence devrait être stipulée dans les documents de planification initiale de l'évaluation et dans la cartographie des parties prenantes élaborée par les évaluateurs et présentée dans les rapports techniques de démarrage et d'évaluation (en annexe si trop détaillée). De cette manière, toutes les parties prenantes essentielles (internes et externes) seraient identifiées et leurs rôles et contributions clairement décrits.

Recommandation 2 : Demander à IDEV de définir les thèmes transversaux à hiérarchiser en se basant sur ce que la politique de la BAD a identifié comme important et de spécifier les thèmes pertinents pour chaque type d'évaluation :

- Les thèmes transversaux pertinents devraient être explicitement identifiés dès le départ (phase de démarrage) et efficacement intégrés de manière transversale, en commençant par la

description du contexte et de l'objet de l'évaluation à travers les objectifs, la portée, les constatations, les conclusions et les recommandations de l'évaluation ;

- Lorsque l'absence de données désagrégées sur ces thèmes empêche leur appréciation complète, l'évaluation peut être l'occasion d'approfondir ces questions en les intégrant dans le cadre d'évaluation et dans les instruments de collecte de données primaires correspondants, et de recommander à la direction de leur accorder plus d'attention dans le suivi des opérations de la Banque.
- En ce qui concerne l'égalité des sexes en particulier, elle ne peut être considérée comme un thème transversal pertinent dans tous les types d'évaluations menées par IDEV, étant donné que la BAD, comme toutes les institutions financières internationales, a une politique qui définit son engagement à « promouvoir l'intégration des questions de genre en tant que moyen pour favoriser la réduction de la pauvreté, le développement économique et l'égalité des sexes sur le continent, en reconnaissant ce dernier point comme un objectif de développement »¹⁶, et puisque IDEV a dirigé la rédaction d'orientations pour l'ECG sur les évaluations axées sur l'égalité des sexes et l'intégration de l'analyse de genre dans l'évaluation au niveau du projet¹⁷, IDEV devrait s'efforcer d'intégrer systématiquement les considérations de genre dans toutes les évaluations impliquant l'appréciation de la qualité de projets ainsi que dans d'autres types d'évaluations pour lesquelles il serait pertinent de le faire.
- S'il a besoin d'un soutien pour parvenir à une intégration efficace de ces questions, IDEV devrait demander une assistance technique spécialisée par le biais d'ateliers spécialisés, de webinaires, etc. visant à renforcer les capacités des chargés de projet dans ces domaines.

Recommandation 3: Veiller à ce que les principes d'évaluation de l'OCDE-CAD et de la BAD-IDEV, relatifs à l'indépendance, à l'impartialité, à l'éthique, au partenariat, à la coordination et au renforcement des capacités soient clairement et explicitement stipulés et expliqués dans les différents produits d'évaluation, de la conception au rapport final, et considérés comme faisant partie intégrante du processus d'évaluation. La preuve de la prise en compte de ces principes pourrait être fournie après la description de la méthodologie d'évaluation, dans l'introduction de divers outils de collecte de données et dans une clause d'exclusion de responsabilité dans les premières pages des rapports d'évaluation (rapports de synthèse et techniques).

Recommandation 4: Prendre acte de la récente décision d'IDEV de remplacer l'astreignante échelle de notation à 6 points par une échelle à 4 points pour les évaluations futures, et envisager de renommer l'intitulé des notes de la nouvelle échelle pour permettre une appréciation plus nuancée entre « Insatisfaisant » et « Satisfaisant » lorsque cela est pertinent ou applicable. Cette échelle devrait permettre de créer une perspective qui reflète plus positivement un certain degré de progrès vers une performance satisfaisante. Nous recommandons de remplacer les notes existantes par les termes « Insatisfaisant », « Partiellement insatisfaisant », « Satisfaisant » et « Très satisfaisant ».

¹⁶ Banque africaine de développement - Fonds africain de développement. Juin 2001. *The Gender Policy (Politique en matière de genre)*. UEDD, p. i.

¹⁷ Voir Groupe de coopération pour l'évaluation. 2017. *Integrating Gender into Project-level Evaluation (Intégrer le genre dans les évaluations de projets)*. Abidjan : BAD-IDEV.

Recommendation 5: Identifier clairement les éléments de base que doit contenir chaque produit d'évaluation et clarifier les circonstances dans lesquelles les TdR, les documents d'orientation et les rapports de démarrage doivent être rédigés.

Recommendation 6: Élaborer des lignes directrices pour les évaluateurs avec des définitions claires des termes clés de l'évaluation, pour bien distinguer le type d'information qui devrait être inclus dans les différentes sections du rapport d'évaluation, par exemple les constatations de l'évaluation, les conclusions, les leçons apprises et les recommandations.

Recommendation 7: S'assurer que les recommandations sont systématiquement débattues avec les sections opérationnelles pertinentes de la Banque et clairement classées par ordre de priorité ou selon un calendrier concordant pour aider à la prise de décision. Tout en évitant la formulation de recommandations trop prescriptives, IDEV devrait identifier les sujets de préoccupation dans les recommandations afin de les relier plus spécifiquement aux résultats des évaluations.

Recommendation 8: Veiller à ce que les annexes du rapport de démarrage comprennent les outils de collecte de données et que les annexes du rapport de synthèse contiennent une matrice d'évaluation abrégée, un outil d'évaluation clé qui relie la portée de l'évaluation à la méthodologie de l'évaluation. La matrice abrégée pourrait inclure les critères d'évaluation et les principales questions correspondantes, ainsi qu'un échantillon d'indicateurs clés (à la fois quantitatifs et qualitatifs, le cas échéant) et les méthodes de collecte de données pour chaque question principale.

Annexe I. Liste des évaluations examinées

ANNÉE	TITRE DE L'ÉVALUATION	TYPOLOGIE
2009	Cameroun : Évaluation de l'assistance de la Banque	Pays
2009	Cap Vert : Évaluation de l'assistance de la Banque 1996-2007	Pays
2009	Évaluation indépendante de la stratégie et du processus de décentralisation à la Banque africaine de développement	Institutionnelle
2010	Évaluation des opérations d'appui aux réformes de la Banque africaine de développement, 1999-2009	Thématique/Sectorielle
2010	Évaluation indépendante de la qualité à l'entrée des opérations et des stratégies FAD 2005-2008	Institutionnelle
2010	Évaluation indépendante de la supervision des projets à la Banque africaine de développement pour la période 2001 - 2008	Institutionnelle
2010	Libération du potentiel des sous-régions de l'Afrique - Revue de l'assistance du Groupe de la Banque aux banques de développement sous-régionales	Thématique/Sectorielle
2011	Gabon - 1996-2008 - Évaluation de l'aide au pays	Pays
2012	Intégration de l'égalité des genres : Une voie vers des résultats ou une voie vers nulle part ?	Autres : Synthèse
2015	Évaluation de l'aide du Groupe de la Banque aux petites et moyennes entreprises (2006-2013)	Thématique/Sectorielle
2015	Évaluation indépendante de l'élaboration et de l'application des politiques et stratégies	Institutionnelle
2016	Évaluation globale des résultats de développement du Groupe de la Banque africaine de développement 2004-2013	Institutionnelle
2016	République démocratique du Congo : Évaluation de la stratégie et du programme de la Banque 2004-2015	Pays
2016	Ghana : Évaluation de la stratégie et du programme de la Banque 2002-2015	Pays
2016	Évaluation d'impact du Programme d'approvisionnement en eau et d'assainissement en milieu rural en Éthiopie 2006-2014	Autres : Impact
2016	Exemple de cas : Afrique du Sud : Évaluation de la stratégie et du programme pays 2004-2015	Pays
2016	Vers une croissance tirée par le secteur privé : Leçons de l'expérience - Rapport de synthèse d'évaluation	Autres : Synthèse
2017	Évaluation des stratégies et programmes de la Banque Africaine de Développement en Côte d'Ivoire 2006-2016	Pays
2017	Évaluation de la stratégie et des opérations d'intégration régionale de la Banque africaine de développement en Afrique centrale, 2011-2016	Autres : Régionale
2017	Stimuler le développement socio-économique local à travers l'électrification rurale	Autres : Groupée

2017	Politique de gestion des ressources humaines et orientations stratégiques de la Banque africaine de développement : Une évaluation formative	Institutionnelle
2017	Appui de la Banque africaine de développement au développement des chaînes de valeur agricoles : Enseignements pour la stratégie « Nourrir l'Afrique »	Thématique/Sectorielle

Annexe II. Modèle d'appréciation de la qualité

Informations sur l'évaluation	
Titre de l'évaluation	
Type de l'évaluation	
Évaluateurs de qualité	
Année de livraison de l'évaluation	
Document(s) disponible(s) et révisé(s)	

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
i) Contexte, objet, finalité, objectifs et portée de l'évaluation		
<p>1. L'historique/le contexte de l'objet de l'évaluation est clairement décrit et inclut tous les facteurs pertinents. <i>Principaux facteurs sociaux, culturels, politiques, économiques et institutionnels qui sont pertinents pour l'objet de l'évaluation.</i></p>		
<p>2. Une description détaillée de l'objet de l'évaluation est incluse. <i>Identification claire de l'objet de l'évaluation, de sa localisation géographique, de son calendrier, de sa finalité, de ses objectifs, des parties prenantes, des bénéficiaires, du budget, du statut de mise en œuvre, des composantes/activités clés, des résultats attendus.</i></p>		
<p>3. La finalité de l'évaluation, y compris son utilisation prévue et les utilisateurs, ainsi que la pertinence de son calendrier, sont clairement présentés. <i>Présentation claire décrivant la finalité de l'évaluation, sa raison</i></p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<i>d'être, l'utilisation prévue, les utilisateurs visés et la pertinence de son calendrier.</i>		
4. Les objectifs spécifiques de l'évaluation sont clairement définis et cohérents avec sa finalité. <i>Liste pertinente d'objectifs d'évaluation spécifiques qui correspondent bien à l'énoncé de la finalité de l'évaluation.</i>		
5. La portée de l'évaluation est claire et cohérente avec la finalité de l'évaluation. <i>Identification claire des zones géographiques et du calendrier, et inclusions/exclusions programmatiques. La portée est alignée avec la finalité de l'évaluation.</i>		
6. Les critères d'évaluation sont identifiés et cohérents avec la finalité de l'évaluation. <i>Référence aux critères d'évaluation standard de l'OCDE-CAD, à savoir : pertinence, efficacité, efficience et durabilité. Si d'autres critères sont utilisés, ils devraient être expliqués.</i>		
7. Les questions transversales sont suffisamment prises en compte dans la description du contexte, de l'objet et de la portée de l'évaluation, le cas échéant. <i>Une attention suffisante est accordée aux questions transversales pertinentes, telles que l'égalité des sexes, l'environnement/le climat, la croissance inclusive, la fragilité, etc., dans la description du contexte, de l'objet et de la portée de l'évaluation, le cas échéant.</i>		
APPRÉCIATION GÉNÉRALE SECTION I	Note moyenne	

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<i>Explication de la note moyenne de la section, résumant les forces et les faiblesses identifiées.</i>	de la section I	
ii) Conception et méthodologie		
<p>8. La conception de l'évaluation est clairement décrite et convenablement alignée avec la finalité de l'évaluation. Dans le cas des évaluations basées sur la théorie, une théorie du changement claire et cohérente de ce qui est évalué, comprenant des hypothèses, est fournie et utilisée pour éclairer les questions d'évaluation.</p> <p><i>Des informations claires et pertinentes sur la conception de l'évaluation, telles que des données expérimentales, quasi-expérimentales, non expérimentales, théoriques, etc. En présence d'une base théorique, description claire et cohérente du modèle/de la théorie du changement de l'objet de l'évaluation. Cela comprend une description des hypothèses sous-jacentes aux liens de causalité et des indicateurs/critères de jugement/points de référence utilisés pour évaluer les résultats.</i></p>		
<p>9. Le rapport identifie clairement les principales sources d'information et explique de manière adéquate les méthodes de collecte, d'échantillonnage et d'analyse des données.</p> <p><i>Identification claire des sources d'information primaires (le cas échéant) et secondaires et explication : des méthodes de collecte des données quantitatives et/ou qualitatives, telles que les sondages en ligne, les entrevues semi-structurées, les groupes de discussion, etc. ; de l'échantillon, y compris des critères de sélection</i></p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<i>pour chaque méthode ; des méthodes d'analyse des données quantitatives et/ou qualitatives, telles que la triangulation, l'analyse statistique, le changement le plus significatif, etc.</i>		
<p>10. Les questions d'évaluation sont bien alignées sur les critères d'évaluation identifiés et reliées à des indicateurs ou critères de jugement clairs et mesurables.</p> <p><i>Les questions d'évaluation sont liées aux critères d'évaluation et présentées dans une matrice d'évaluation qui doit comporter au moins 4 colonnes (critères/questions, sous-questions, indicateurs et méthodes de collecte de données pour chaque question) et figurer en annexe ou dans le rapport.</i></p>		
<p>11. Les risques et les limites de l'évaluation, ainsi que les stratégies d'atténuation appropriées pour les maîtriser, sont clairement expliqués.</p> <p><i>Description claire des risques et des limites de l'évaluation rencontrés et des stratégies d'atténuation utilisées. Toute différence significative par rapport à l'approche prévue doit être expliquée.</i></p>		
<p>APPRÉCIATION GÉNÉRALE SECTION II</p> <p><i>Explication de la note moyenne de la section, résumant les forces et les faiblesses identifiées.</i></p>	<p>Note moyenne de la section II</p>	
iii) Processus		
<p>12. Il existe des liens logiques entre les informations présentées dans les principaux produits d'évaluation, par exemple les termes de référence, le document d'orientation, le rapport de démarrage et le rapport de synthèse.</p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<p><i>Cohérence dans les objectifs, la portée, la conception et la méthodologie de l'évaluation, avec chaque produit successif s'appuyant sur le précédent. Toute différence d'approche introduite au cours du processus doit être expliquée.</i></p>		
<p>13. Le rapport identifie les principales parties prenantes impliquées dans le processus d'évaluation et explique comment s'opère leur participation. <i>Identification des principales parties prenantes consultées pour l'évaluation, de leur niveau de consultation, des rôles et responsabilités de chaque intervenant, y compris d'un comité directeur ou d'un autre comité consultatif, le cas échéant.</i></p>		
<p>14. Le processus d'assurance-qualité de l'évaluation est clairement décrit. <i>Des informations claires sur la façon dont la qualité a été/pourrait être assurée dans l'évaluation, par exemple, la participation des parties prenantes dans la validation des résultats, l'examen par les pairs du rapport d'évaluation, etc.</i></p>		
<p>15. Des éléments de preuve indiquent que les résultats de l'évaluation ont été/pourront être diffusés/communiqués. <i>Plan de partage des connaissances dans le document d'orientation et/ou preuve que les principales constatations de l'évaluation ont été diffusées/communiquées lors d'un débriefing, d'un atelier, dans une publication, etc.</i></p>		
<p>16. Il existe des preuves de la prise en compte d'autres principes d'évaluation de l'OCDE-CAD et de la BAD-IDEV, à savoir l'indépendance,</p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<p>l'impartialité, l'éthique, la coordination et l'alignement, le partenariat et le renforcement des capacités.</p> <p><i>Preuve : de l'indépendance organisationnelle et de la prévention des conflits d'intérêts ; de l'absence de biais dans les constatations répertoriées ; du suivi des directives éthiques et des codes de conduite ; de la prise en compte des plans, activités et politiques d'évaluation nationaux et locaux ; et du soutien et du renforcement des leçons apprises de l'évaluation chez les membres régionaux.</i></p>		
<p>APPRÉCIATION GÉNÉRALE SECTION III</p> <p><i>Explication de la note moyenne de la section, résumant les forces et les faiblesses identifiées.</i></p>	<p>Note moyenne de la section III</p>	
<p>iv) Clarté et fiabilité des rapports</p>		
<p>17. Le rapport d'évaluation a une structure claire et logique et un résumé analytique concis et représentatif.</p> <p><i>Ordre logique de présentation : historique/contexte, objet de l'évaluation, finalité de l'évaluation, objectifs, portée, méthodologie, constatations, etc. Résumé analytique succinct couvrant tous les éléments clés de l'évaluation, 5 pages maximum par exemple.</i></p>		
<p>18. Le rapport d'évaluation présente une analyse claire et les constatations reposent sur plusieurs sources de données (plus de 2).</p> <p><i>Preuve de triangulation (au moins deux lignes de preuve présentées dans les constatations) et diversité des perspectives. Preuve que les lacunes en termes de données, s'il y en a, sont abordées et, le cas échéant, que des mises en garde</i></p>		

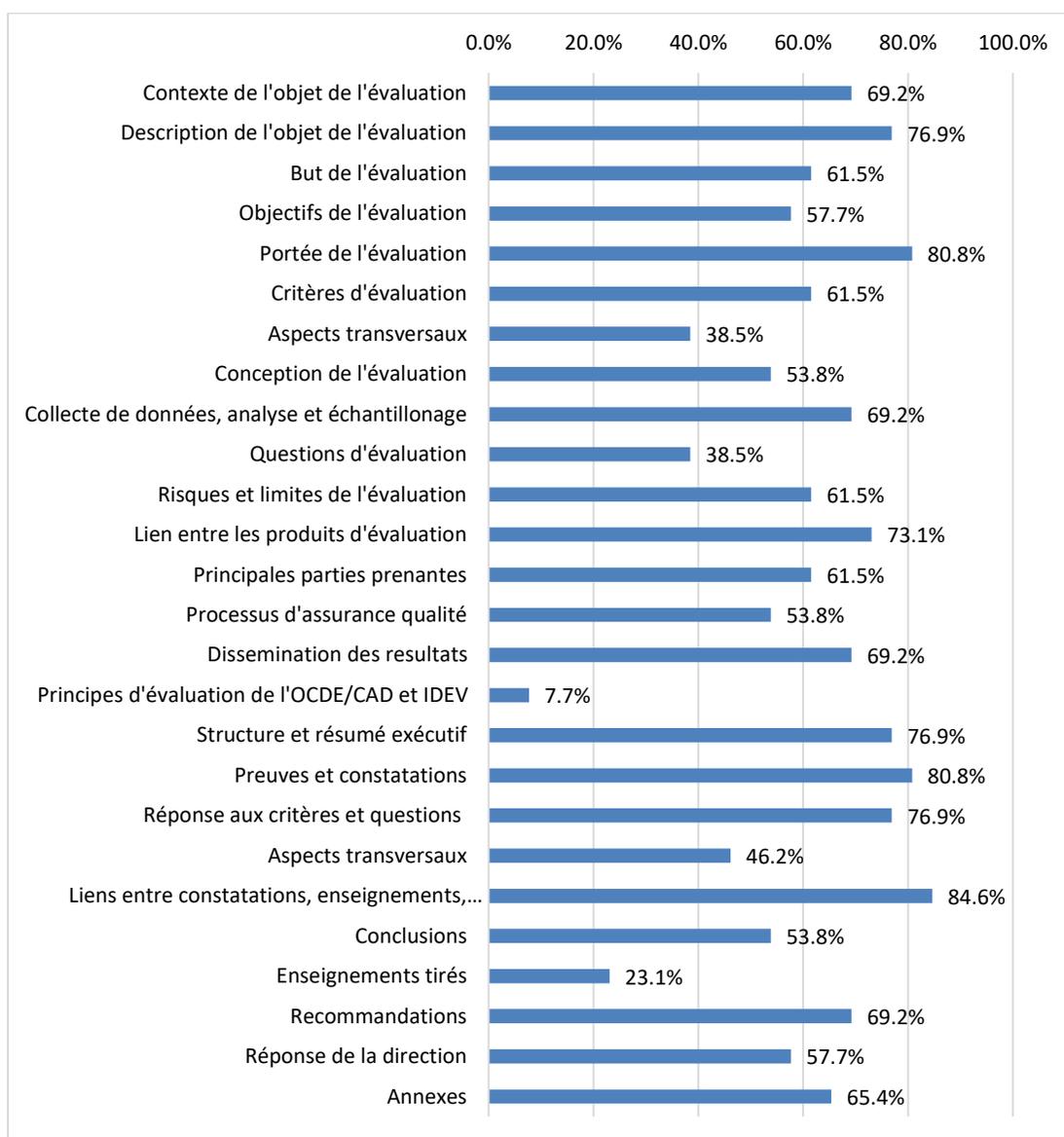
CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<i>sont fournies pour guider le lecteur sur l'interprétation des constatations.</i>		
<p>19. Les critères d'évaluation sont respectés et les questions trouvent systématiquement une réponse.</p> <p><i>Les constatations devraient systématiquement répondre aux critères d'évaluation et aux questions et explications fournies pour les interrogations auxquelles il n'était pas possible de répondre ; toute révision des questions originales est également documentée.</i></p>		
<p>20. Les questions transversales sont suffisamment évaluées dans les constatations et les conclusions, ainsi que dans les recommandations (le cas échéant).</p> <p><i>Évaluation des questions transversales, telles que l'égalité des sexes, l'environnement/le climat, la croissance inclusive, la fragilité, etc., à travers les constatations et conclusions de l'évaluation, et les recommandations (le cas échéant).</i></p>		
<p>21. Les liens entre les éléments de preuve, les constatations et les conclusions, les leçons apprises (le cas échéant) et les recommandations sont évidents.</p> <p><i>Cohérence entre les constatations, les conclusions, les leçons apprises (le cas échéant) et les recommandations. Les conclusions, les leçons apprises et les recommandations ne devraient pas introduire de nouvelles informations non abordées dans les constatations.</i></p>		
<p>22. Les conclusions ajoutent de la valeur aux constatations.</p> <p><i>Conclusions qui fournissent certaines connaissances de l'évaluateur basées sur les constatations liés aux principaux critères d'évaluation, par exemple, la description des implications prévisibles des</i></p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<p><i>constatations pour les futures initiatives. Les conclusions devraient inclure une description des principales forces et faiblesses de l'objet de l'évaluation, mais ne devraient pas simplement répéter ou résumer les principales constatations.</i></p>		
<p>23. Les leçons apprises sont clairement et correctement identifiées.</p> <p><i>Leçons qui sont clairement énoncées et qui démontrent une plus grande pertinence/applicabilité allant au-delà de l'objet de l'évaluation elle-même.</i></p>		
<p>24. Les recommandations sont claires et précises, mais non prescriptives, tout en indiquant un calendrier.</p> <p><i>Recommandations qui sont organisées par ordre de priorité et exécutables tout en n'étant pas prescriptives.</i></p>		
<p>25. Il existe des preuves d'une réponse adéquate des gestionnaires aux recommandations de l'évaluation (critère non applicable aux évaluations de synthèse et groupées)</p> <p><i>Type de réponse de la direction, par exemple, acceptation de la plupart ou de la totalité des recommandations et du calendrier de mise en œuvre, explication du rejet d'une recommandation.</i></p>		

CRITÈRES/SOUS-CRITÈRES	NOTATION	COMMENTAIRES DE L'ÉVALUATEUR*
<p>26. Le rapport (technique et/ou de synthèse) contient les annexes pertinentes.</p> <p><i>Informations complémentaires, telles que la matrice d'évaluation, liste des parties prenantes consultées, autres sources d'information (à condition qu'elles ne soient pas en conflit avec le respect de la vie privée et la confidentialité), outils de collecte de données et tableaux de données/preuves supplémentaires ou contexte pour soutenir l'analyse et les constatations présentées.</i></p>		
<p>APPRÉCIATION GÉNÉRALE SECTION IV</p> <p><i>Explication de la note moyenne de la section, résumant les forces et les faiblesses identifiées.</i></p>	<p>Note moyenne de la section IV</p>	
<p>APPRÉCIATION GÉNÉRALE DE L'ÉVALUATION</p> <p><i>Résumé des principales forces et faiblesses de l'évaluation.</i></p>	<p>Pourcentage de critères notés 3 et 4</p>	

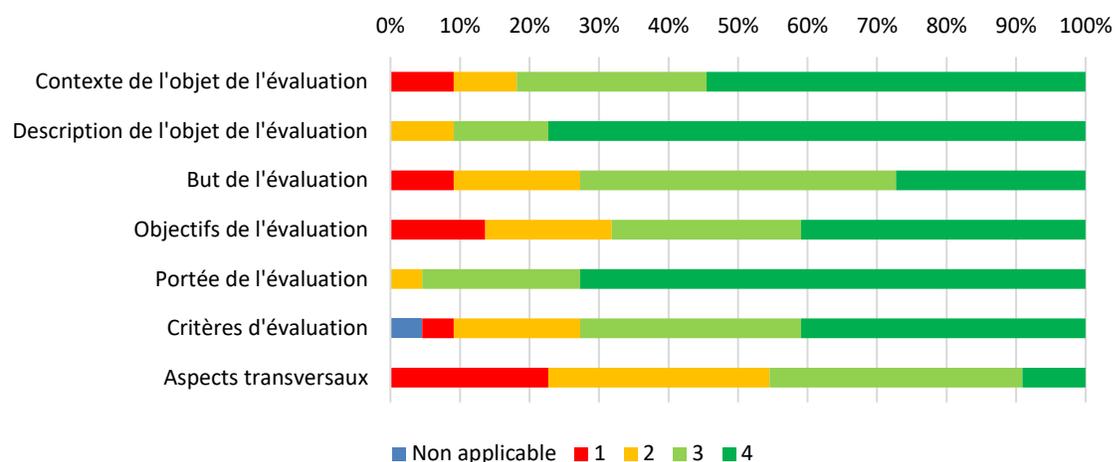
*Les commentaires de l'évaluateur s'appuieront principalement sur les informations contenues dans le rapport de synthèse de l'évaluation mais tous les documents associés seront examinés, tels que les rapports techniques détaillés et les rapports de démarrage si disponibles, afin de vérifier la cohérence méthodologique par exemple, mais aussi pour rechercher des informations qui peuvent ne pas se trouver dans le rapport de synthèse. Lorsque l'information n'a pas été trouvée dans le rapport de synthèse mais dans d'autres documents associés, ou qu'elle ne figure dans aucun des documents examinés, cela doit être explicitement mentionné.

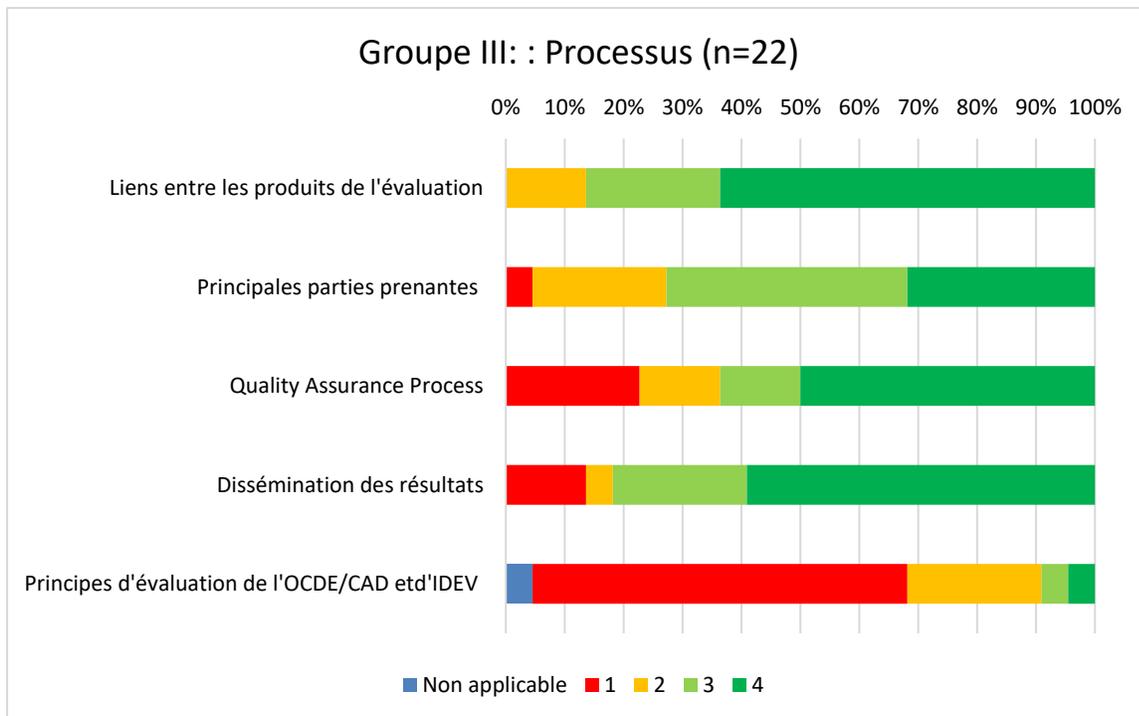
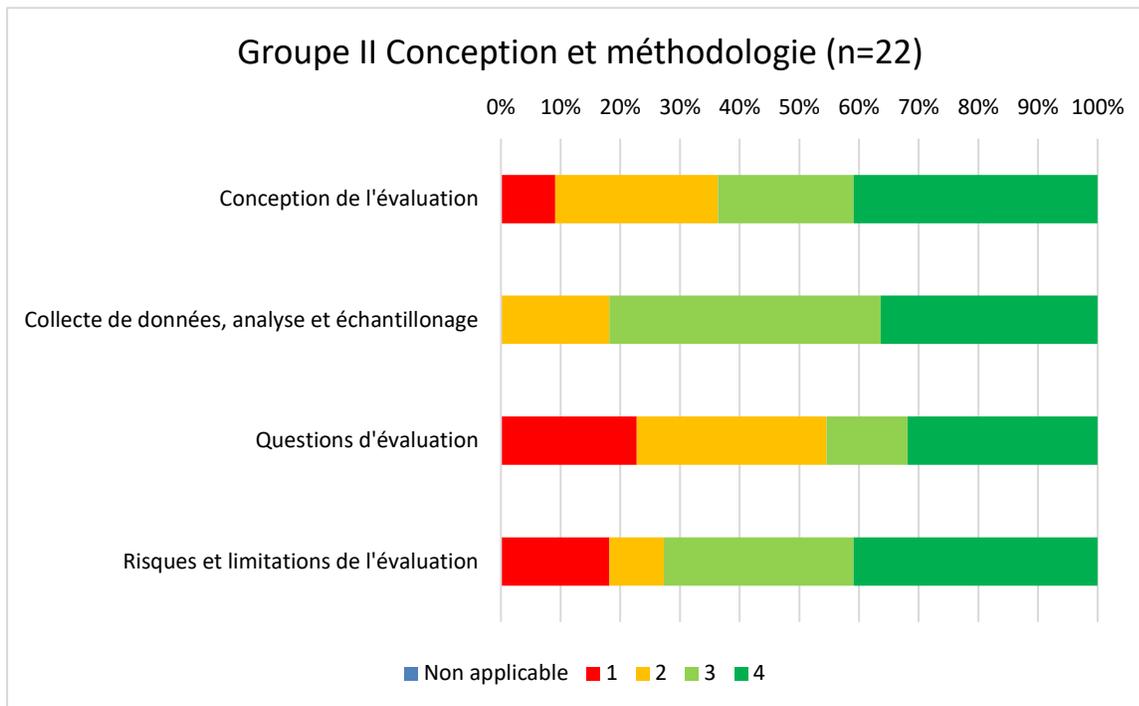
Annexe III. Pourcentage de critères jugés satisfaisants et très satisfaisants, toutes évaluations confondues

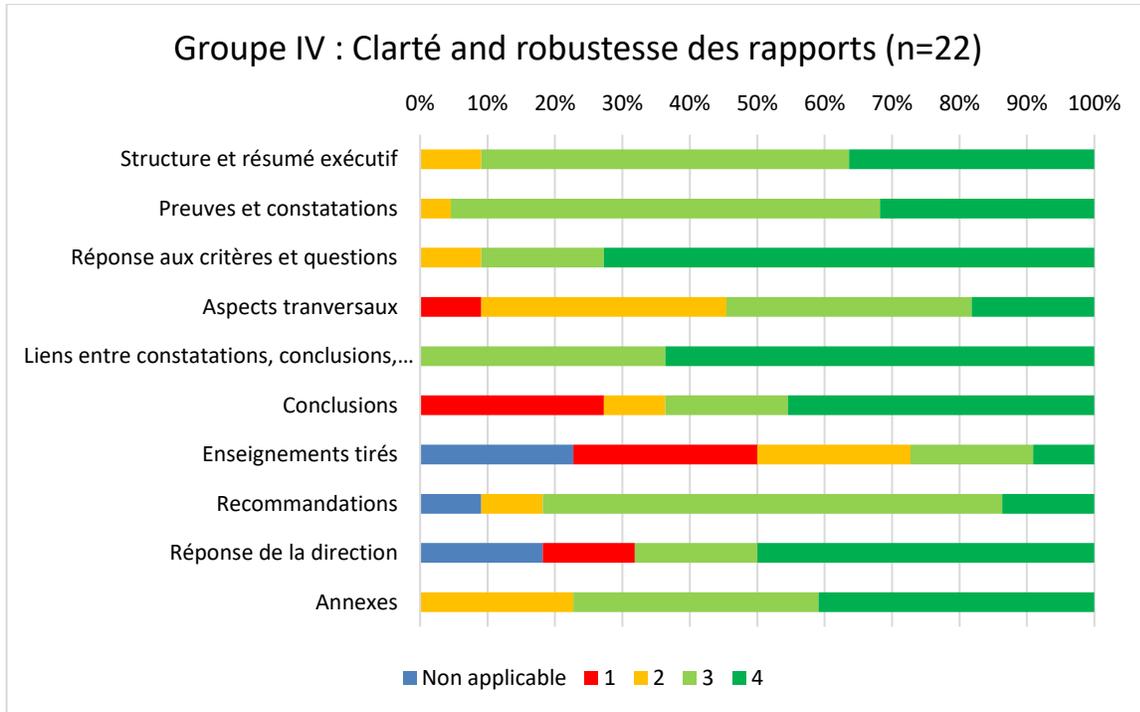


Annexe IV. Résultats de l'appréciation de la qualité par critère, toutes évaluations confondues

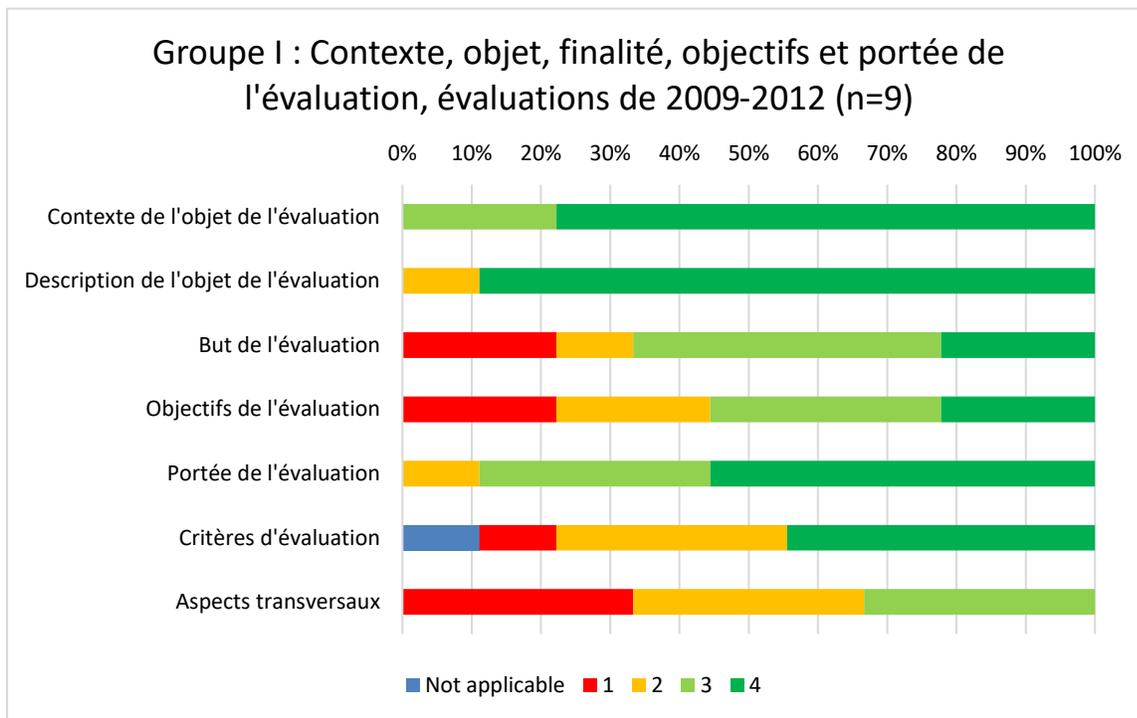
Groupe I : Contexte, objet, finalité, objectifs et portée de l'évaluation (n=22)

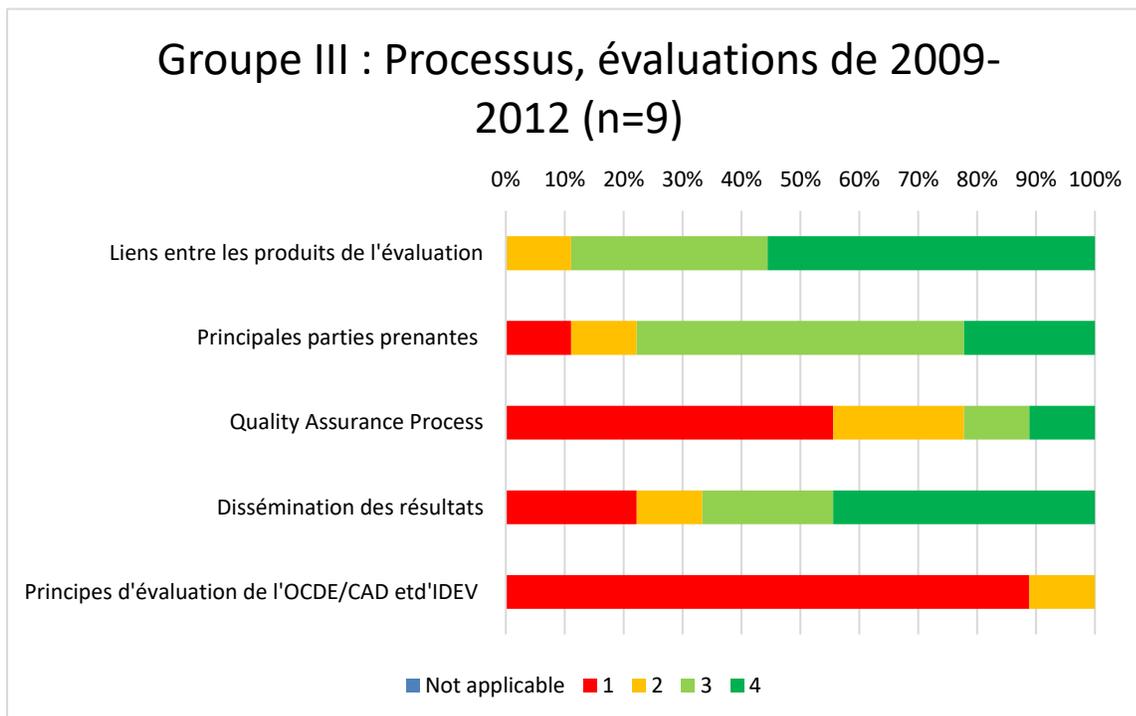
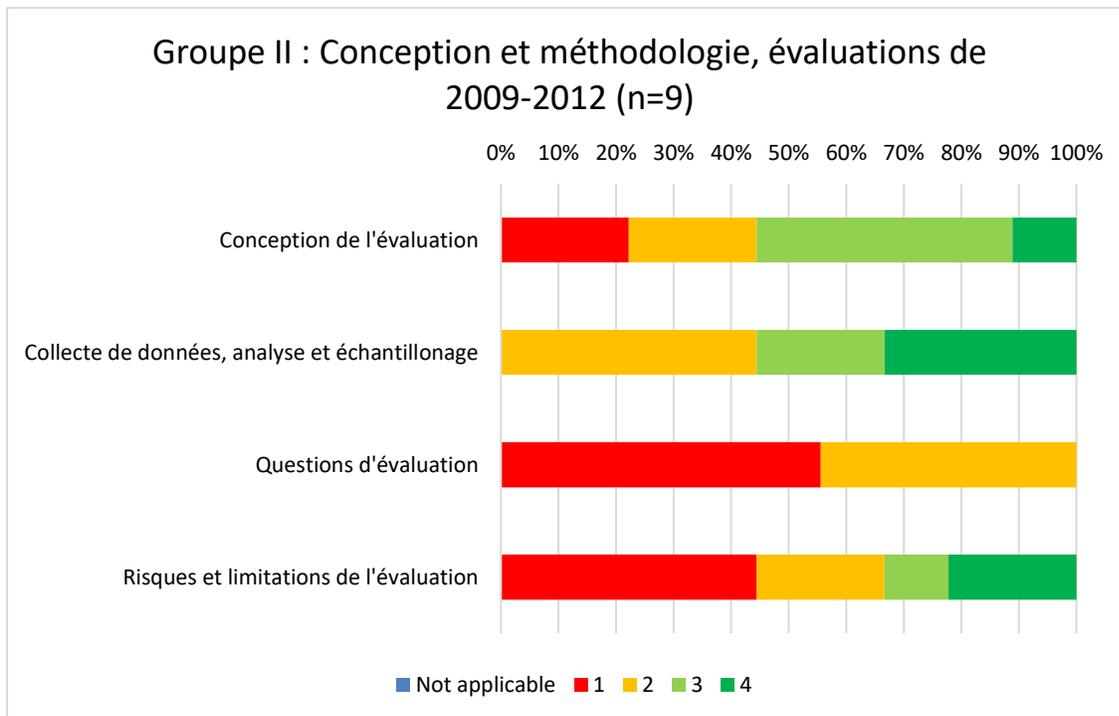


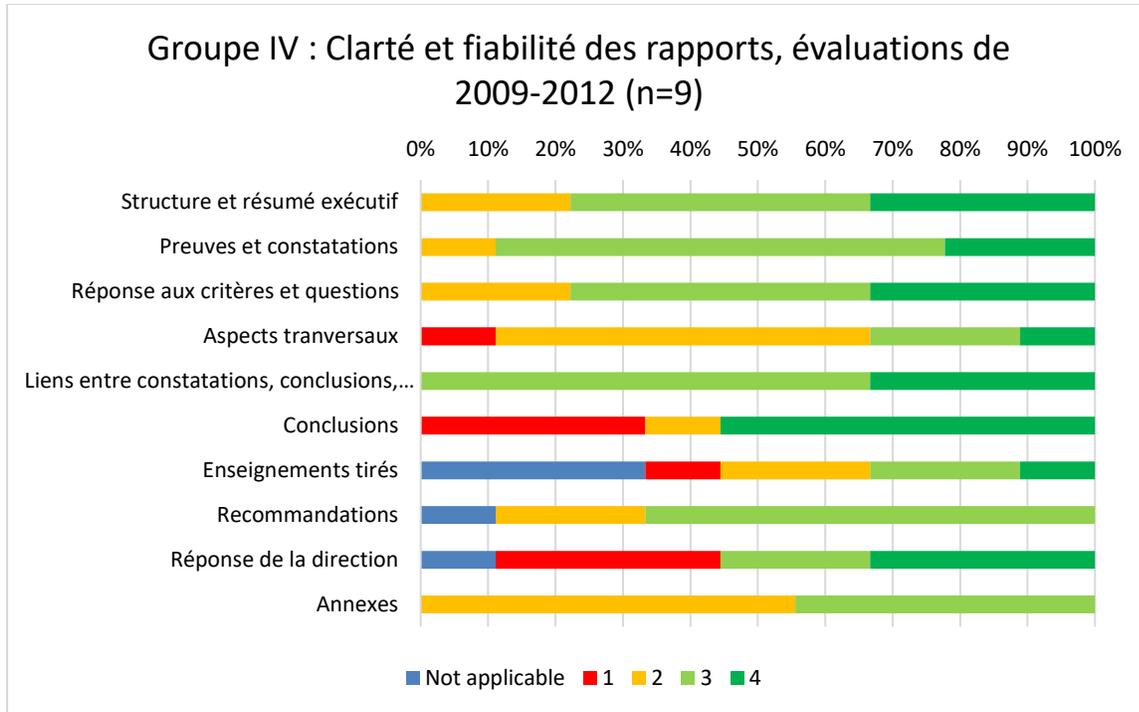




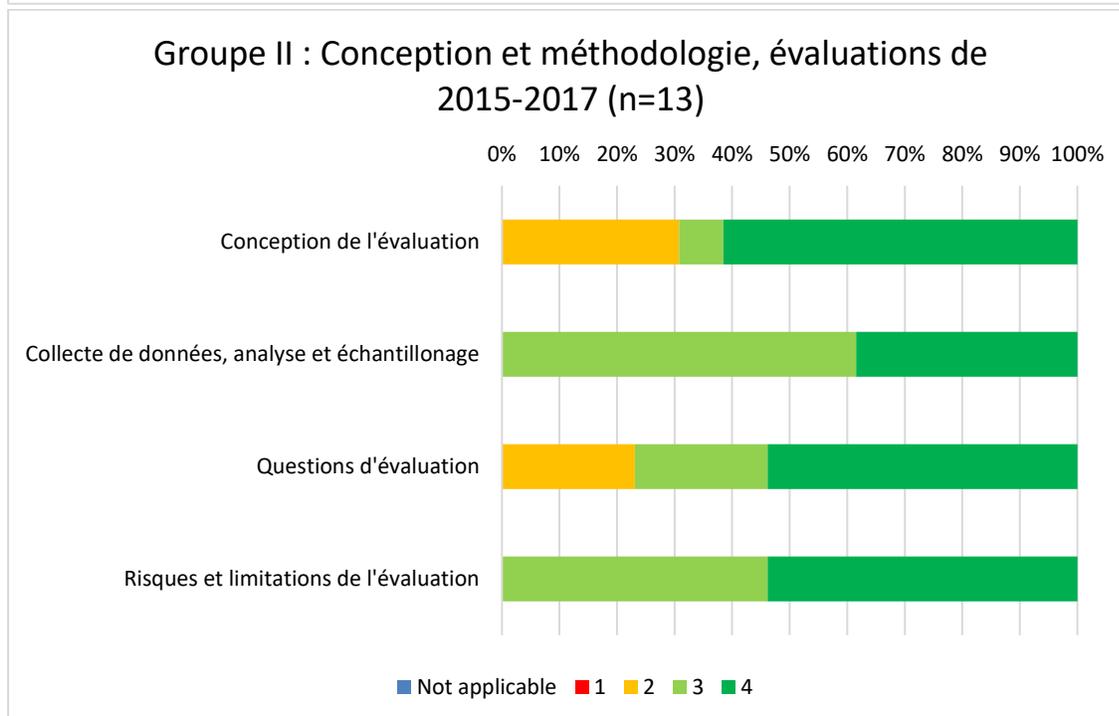
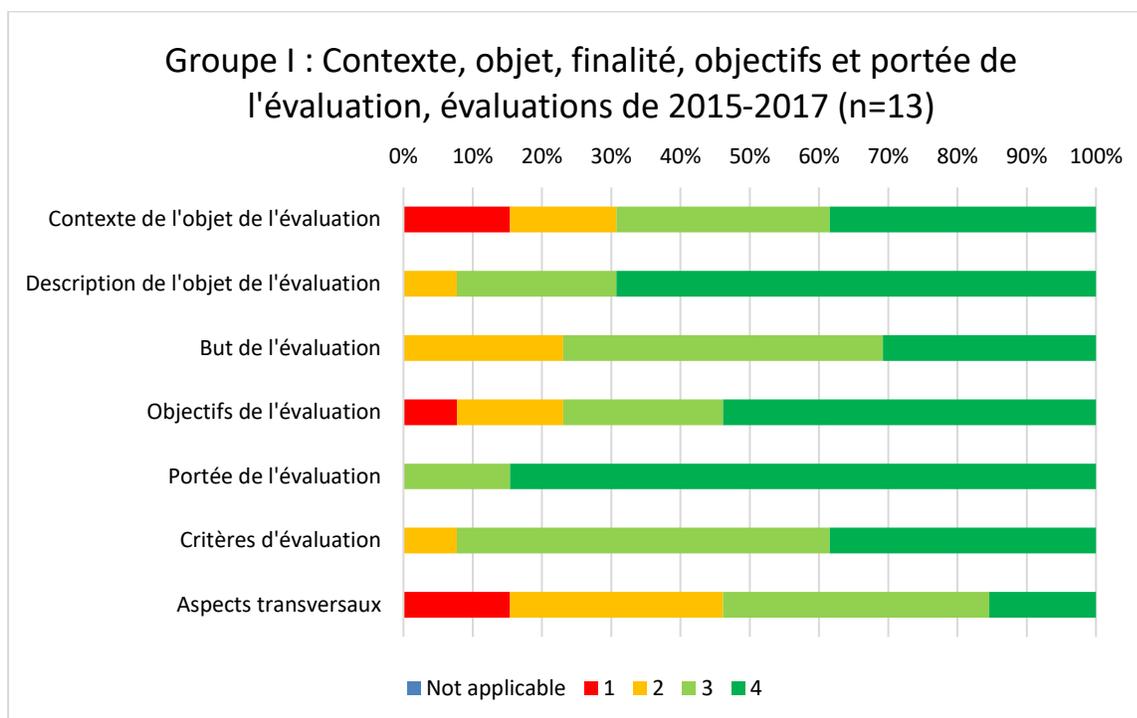
Annexe V. Résultats de l'appréciation de la qualité par critère, évaluations de la première période (2009-2012)



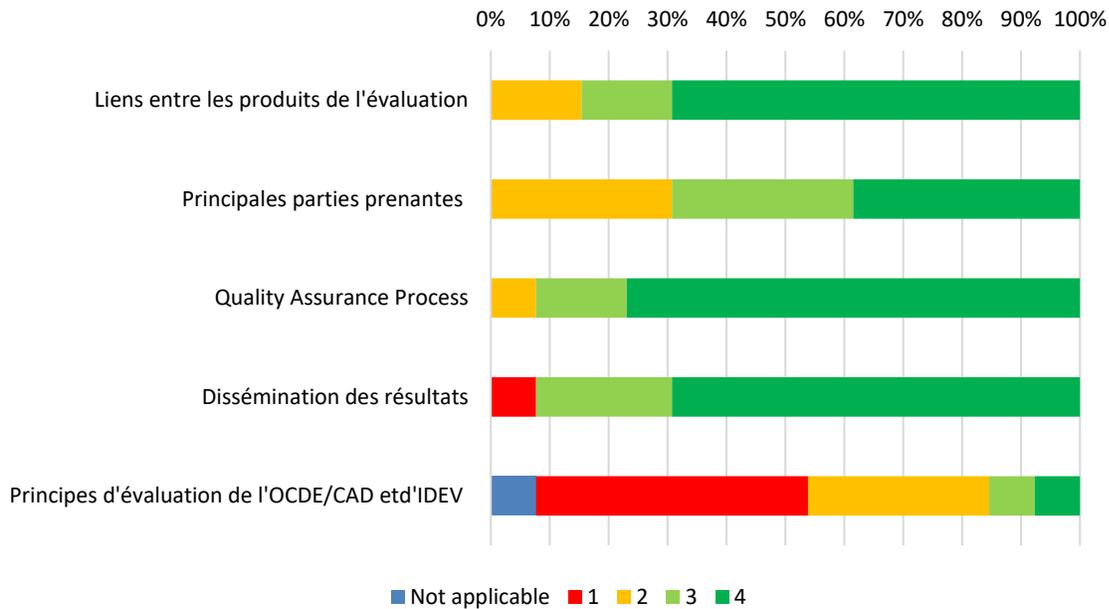




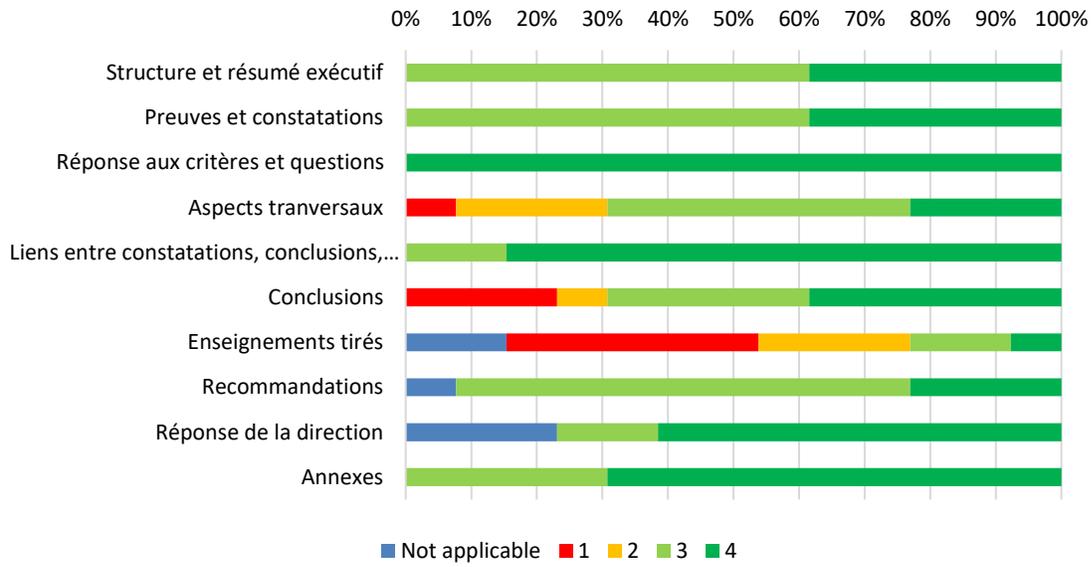
Annexe VI. Résultats de l'appréciation de la qualité par critère, évaluations de la deuxième période (2015-2017)



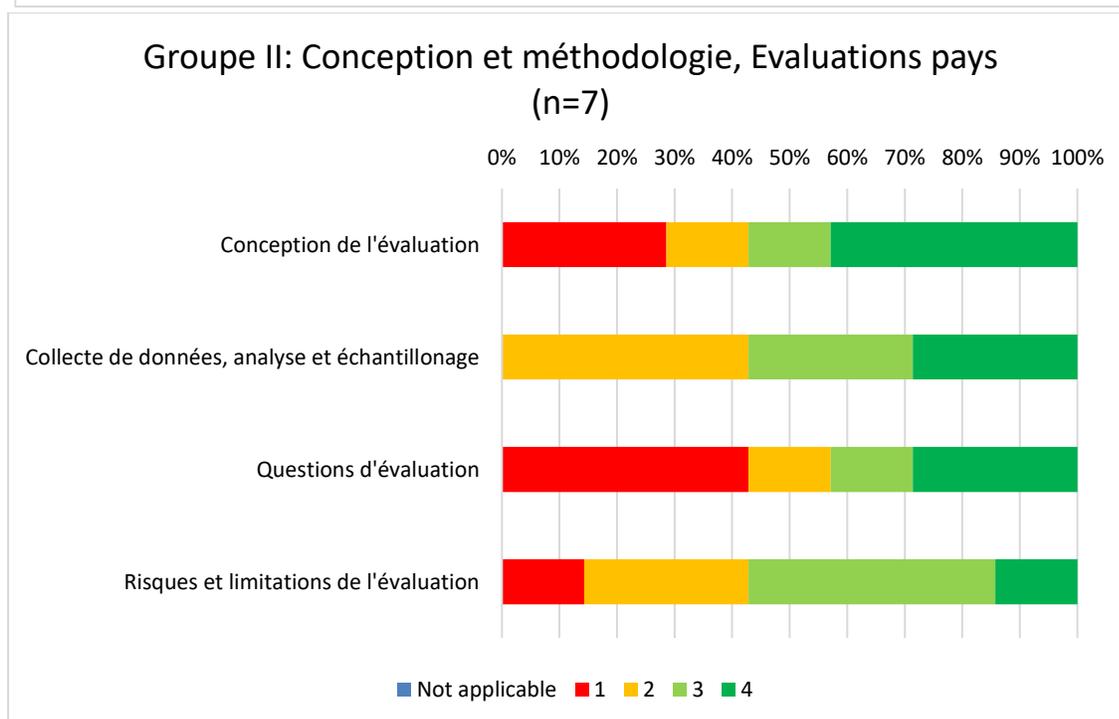
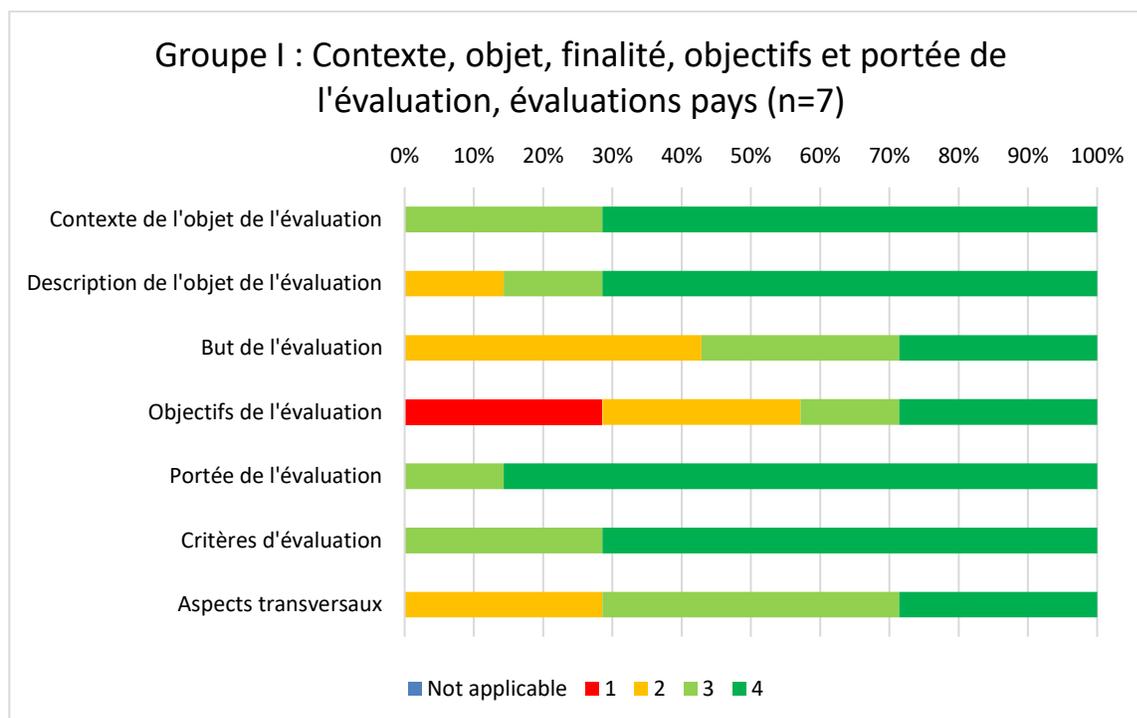
Groupe III : Processus, évaluations de 2015-2017 (n=13)

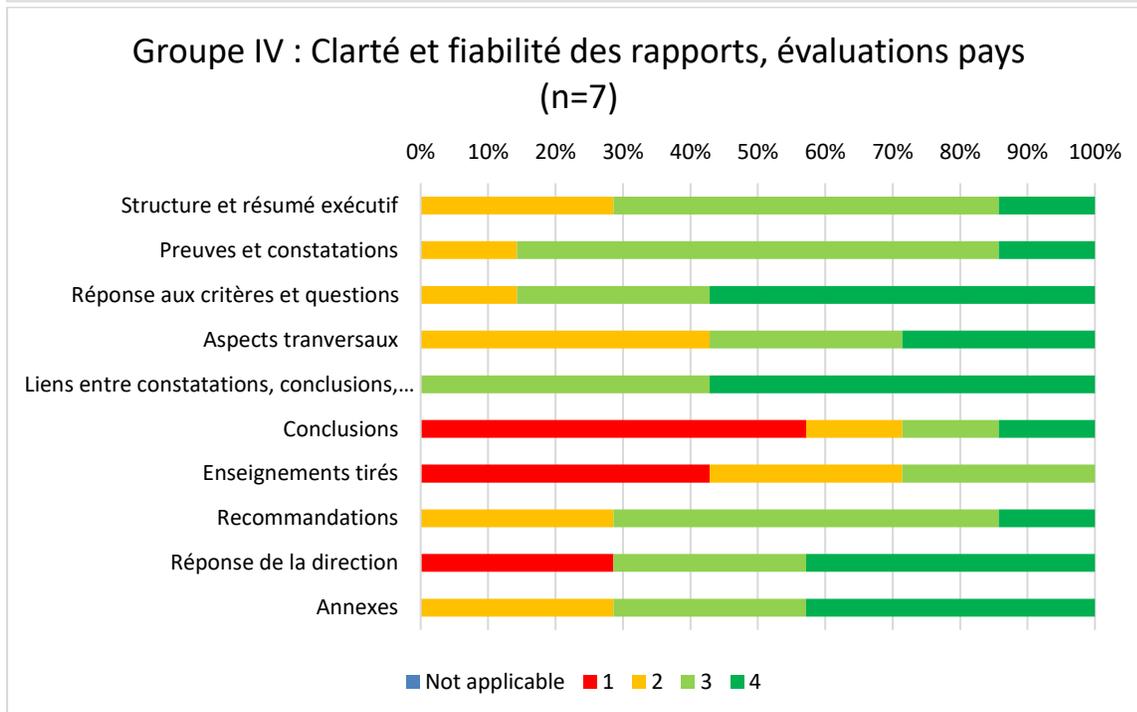
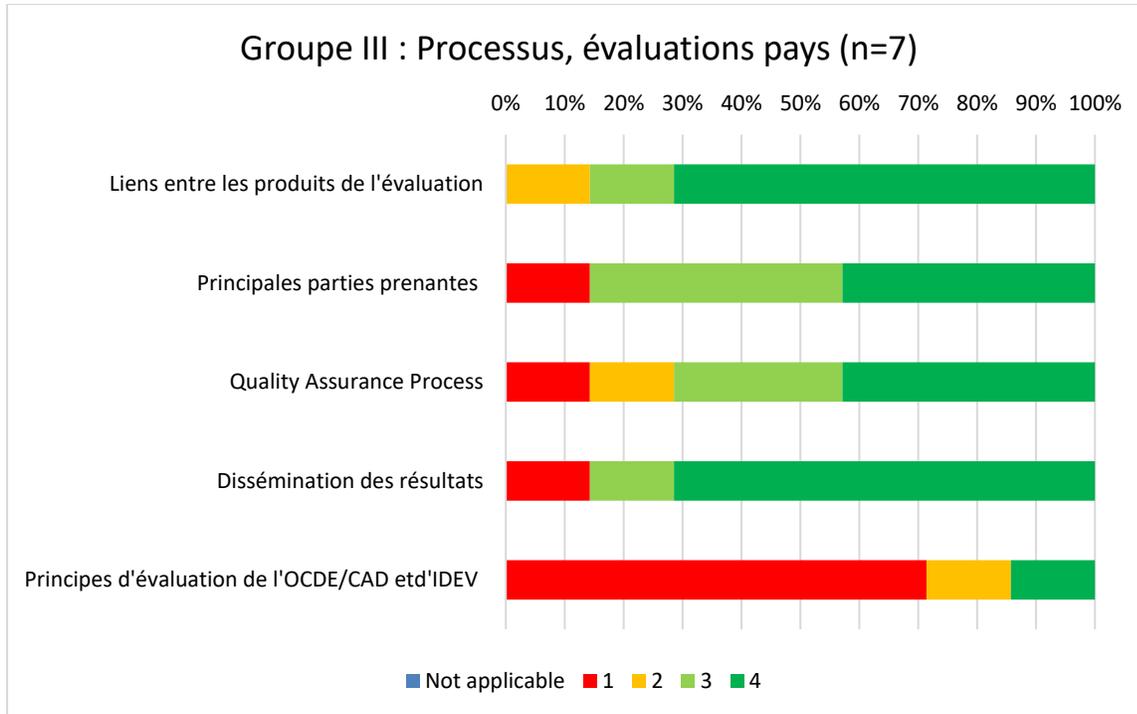


Groupe IV : Clarté et fiabilité des rapports, évaluations de 2015-2017 (n=13)

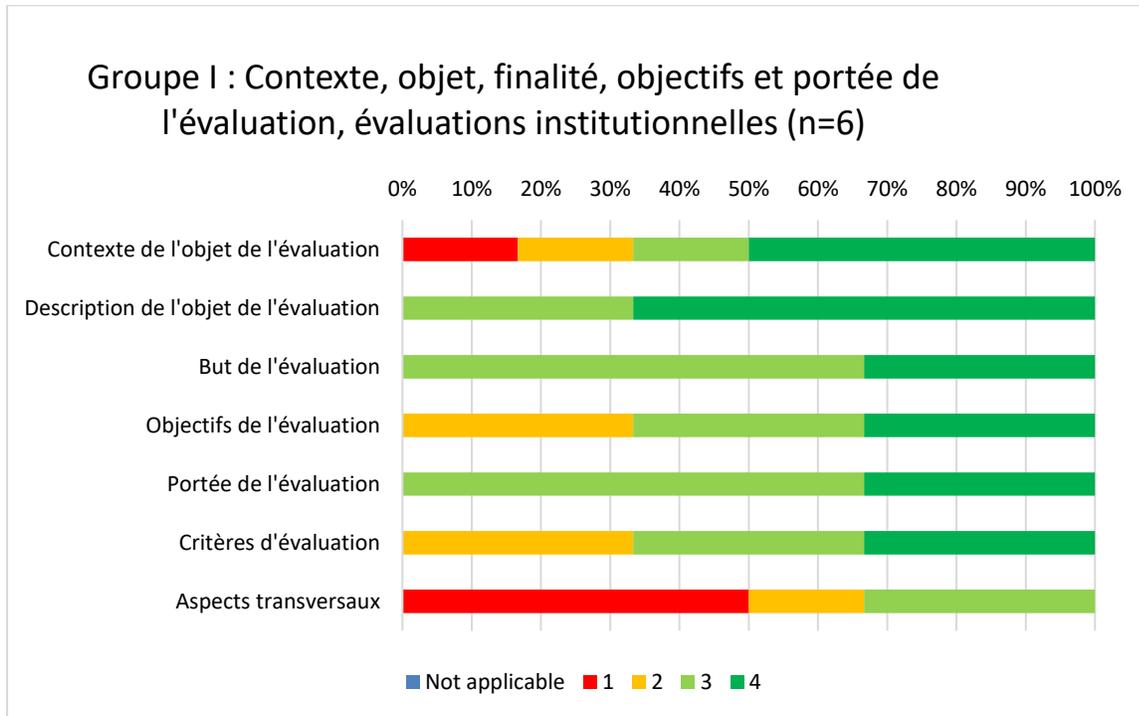


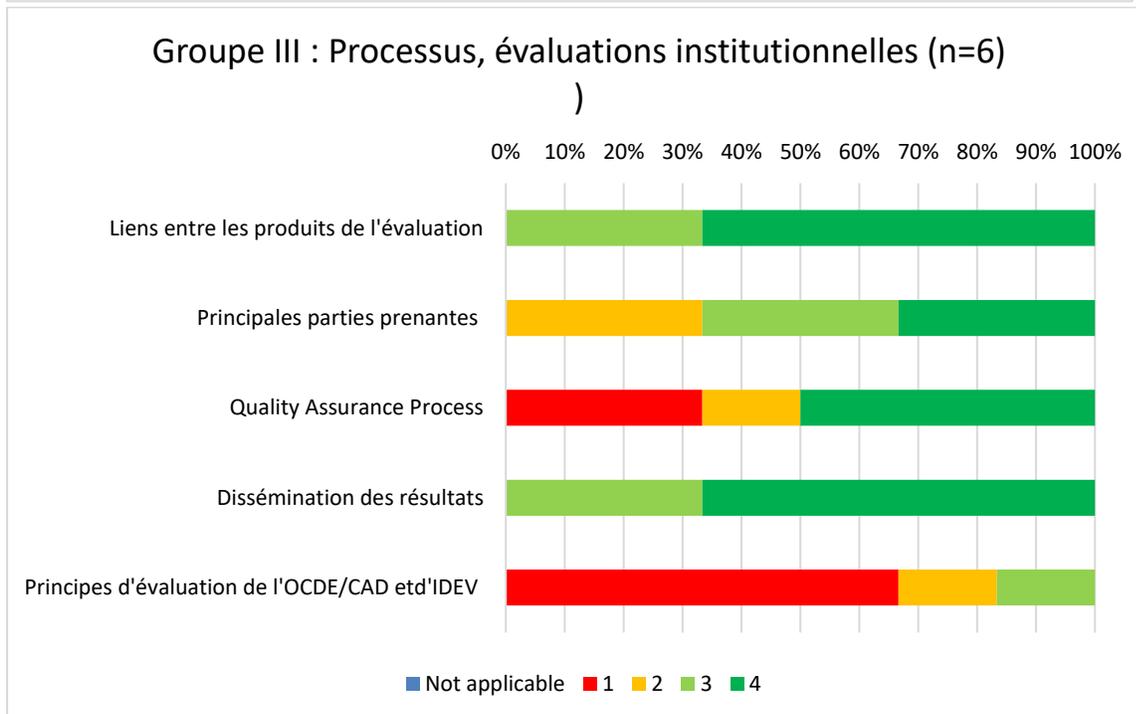
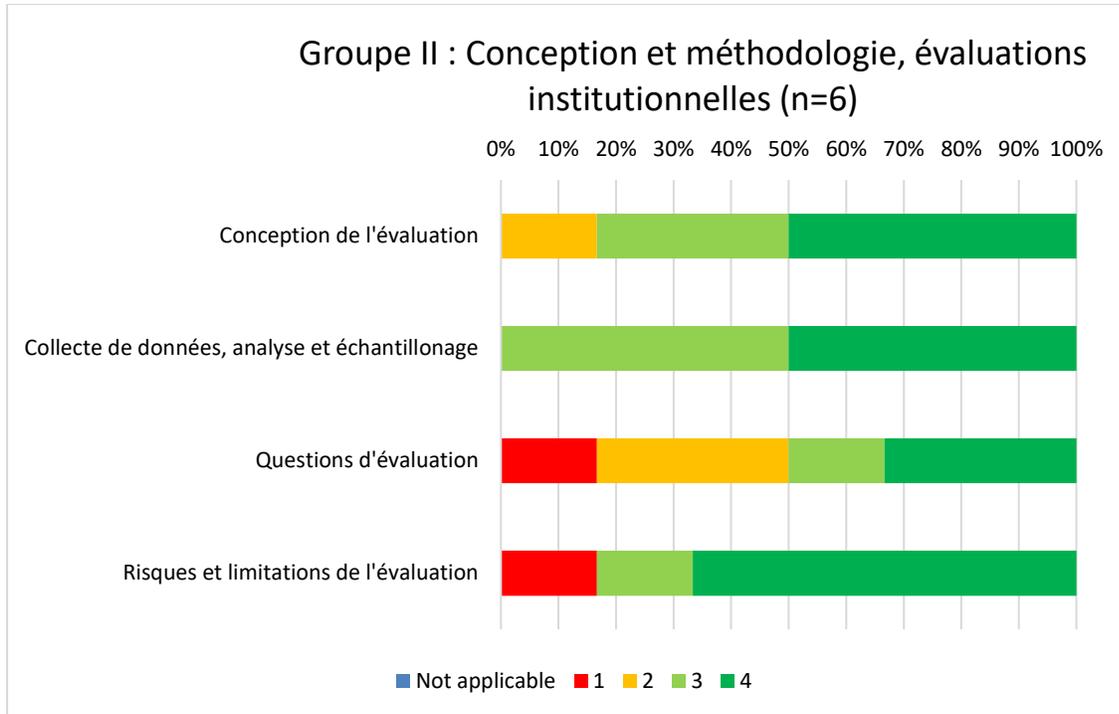
Annexe VII. Résultats de l'appréciation de la qualité par critère, évaluations par pays

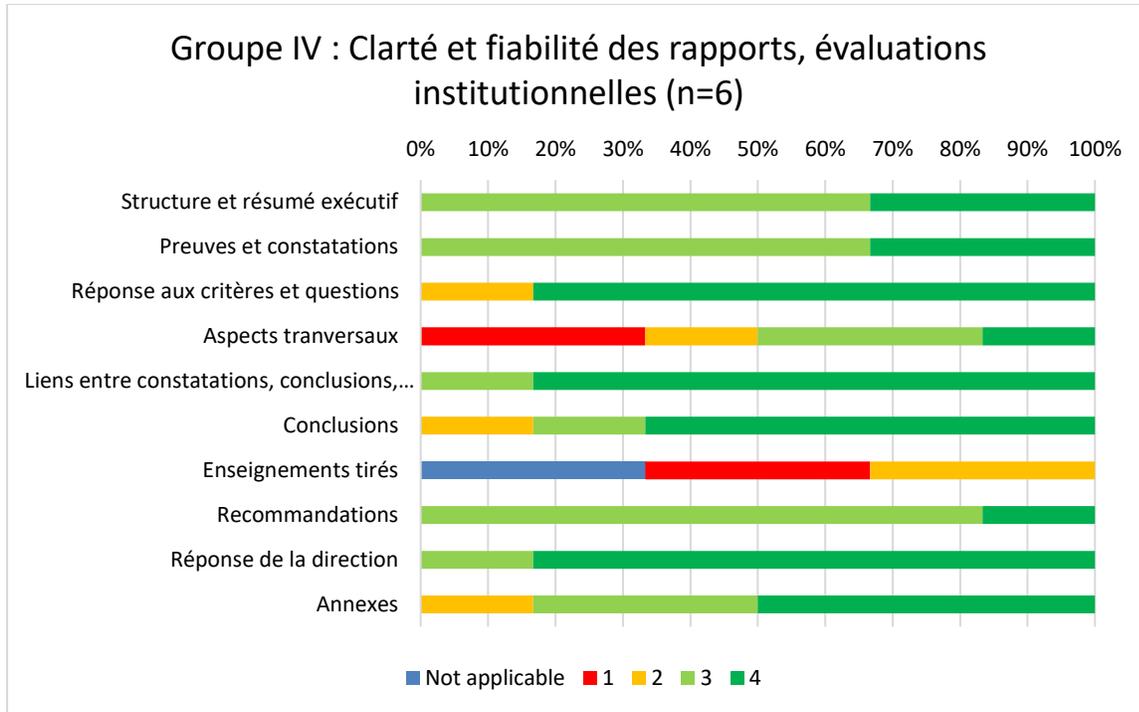




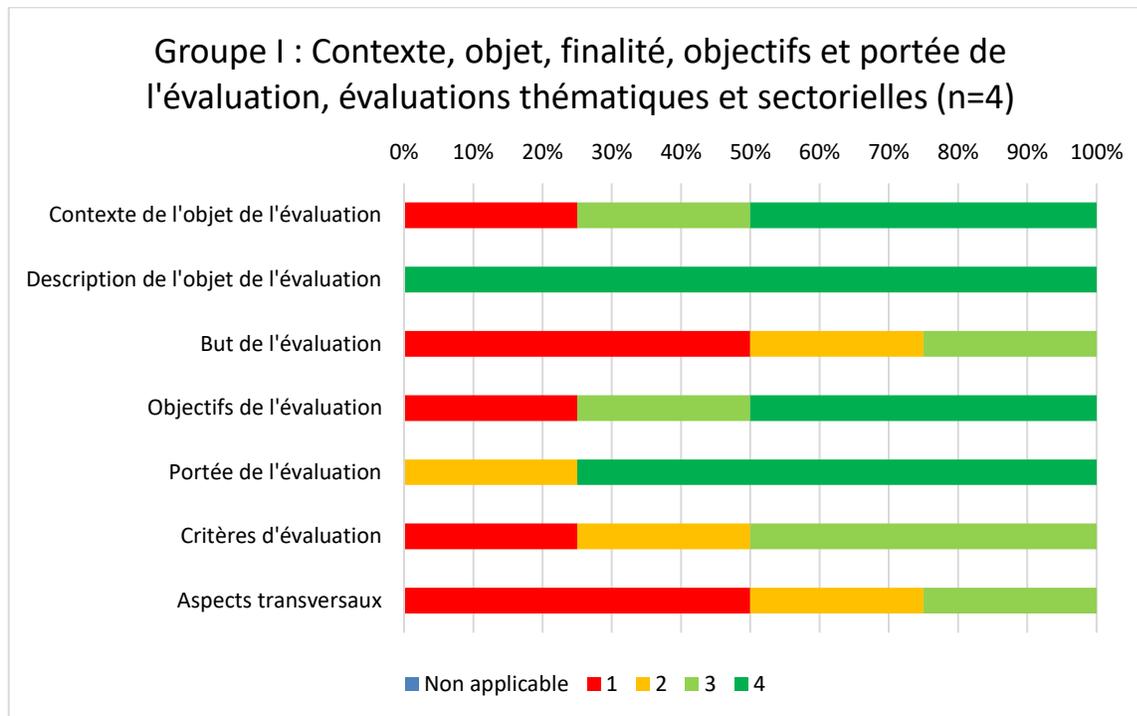
Annexe VIII. Résultats de l'appréciation de la qualité par critère, évaluations institutionnelles



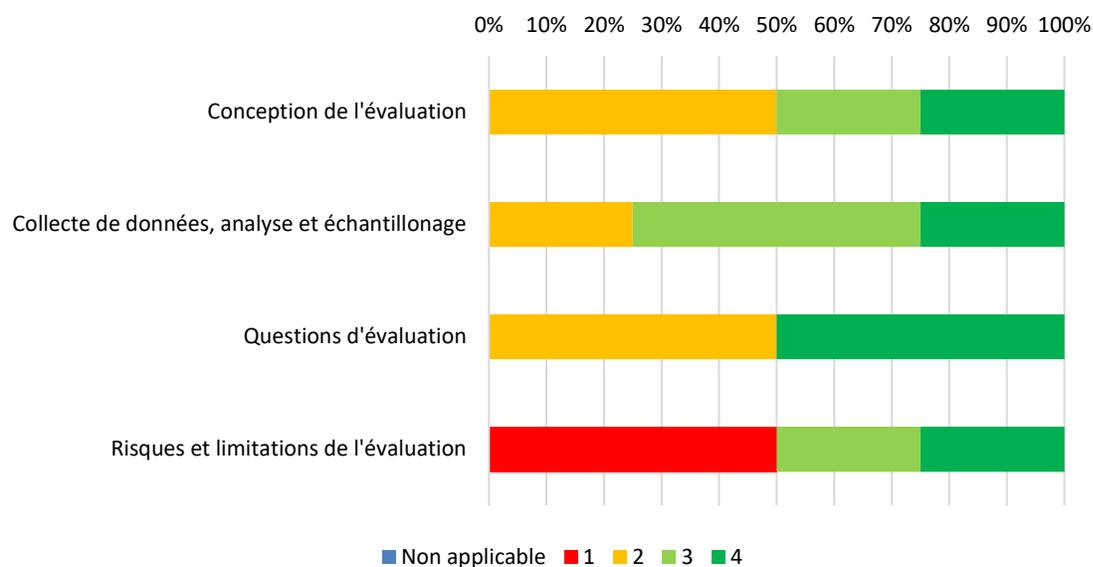




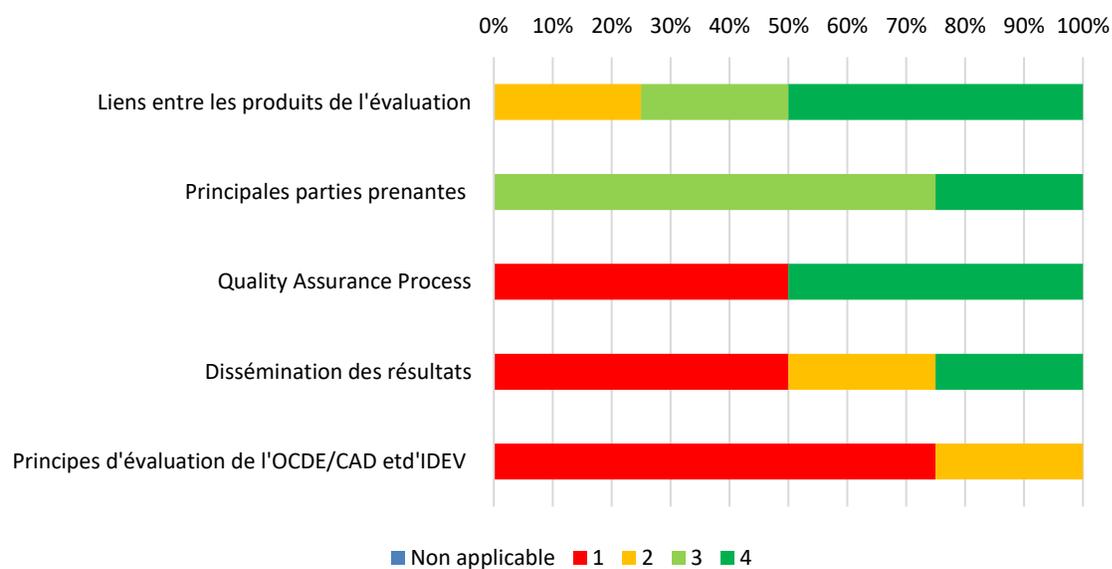
Annexe IX. Résultats de l'appréciation de la qualité par critère, évaluations thématiques et sectorielles

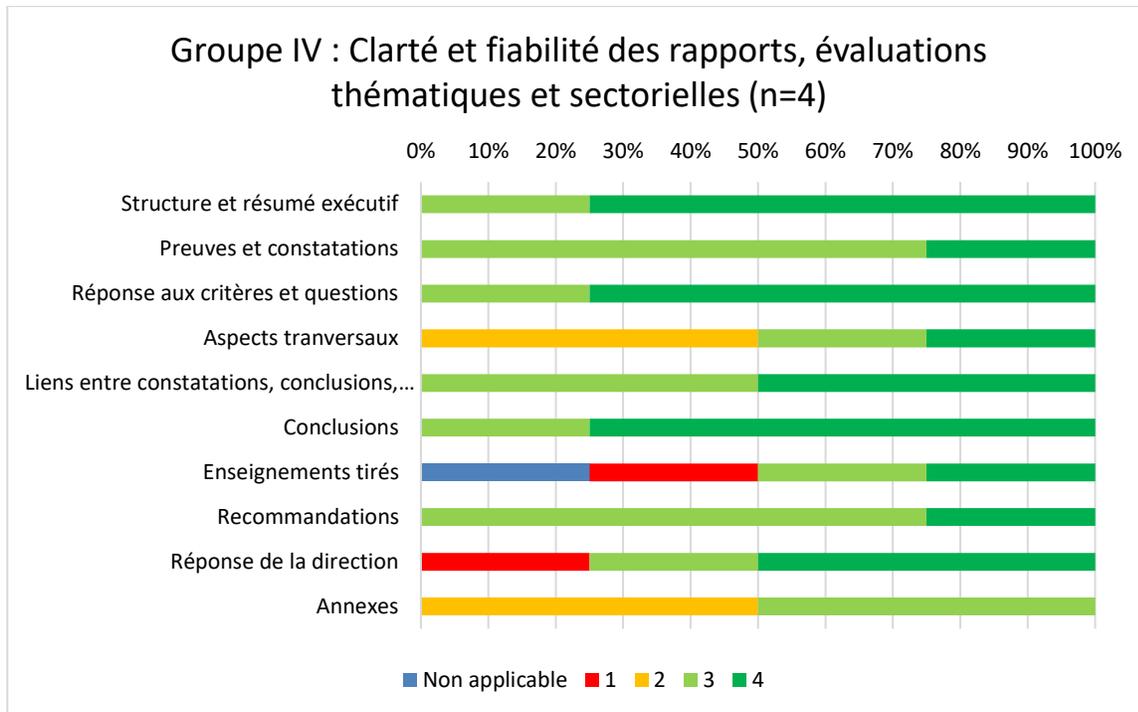


Groupe II : Conception et méthodologie, évaluations thématiques et sectorielles (n=4)

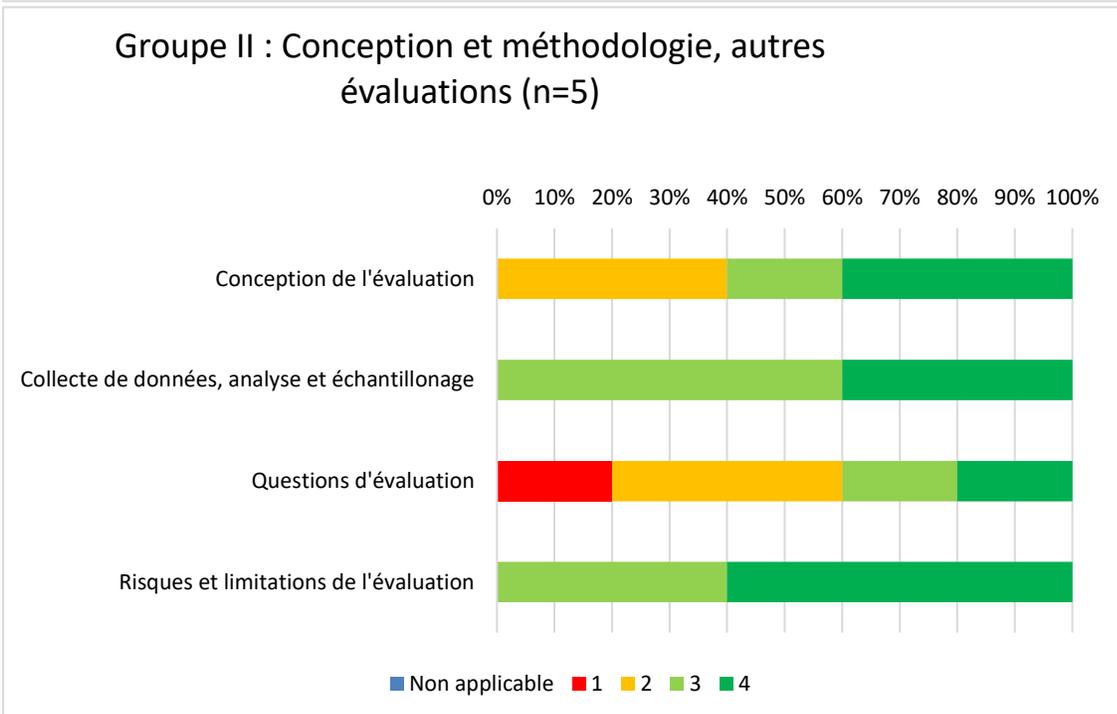
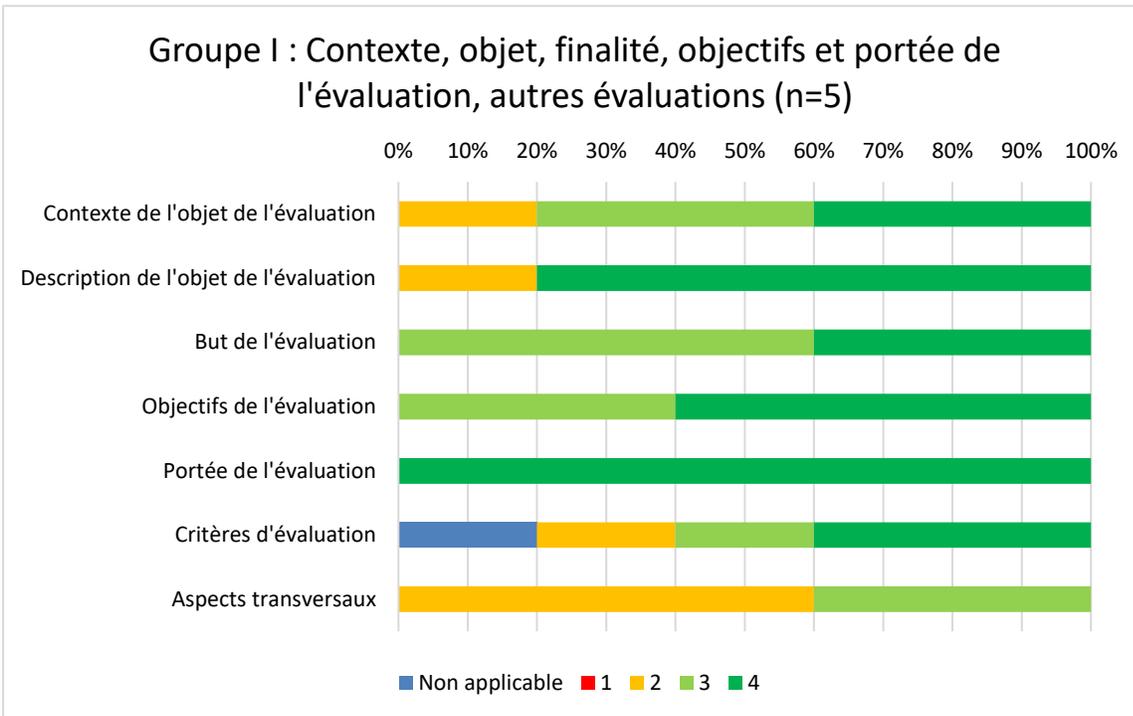


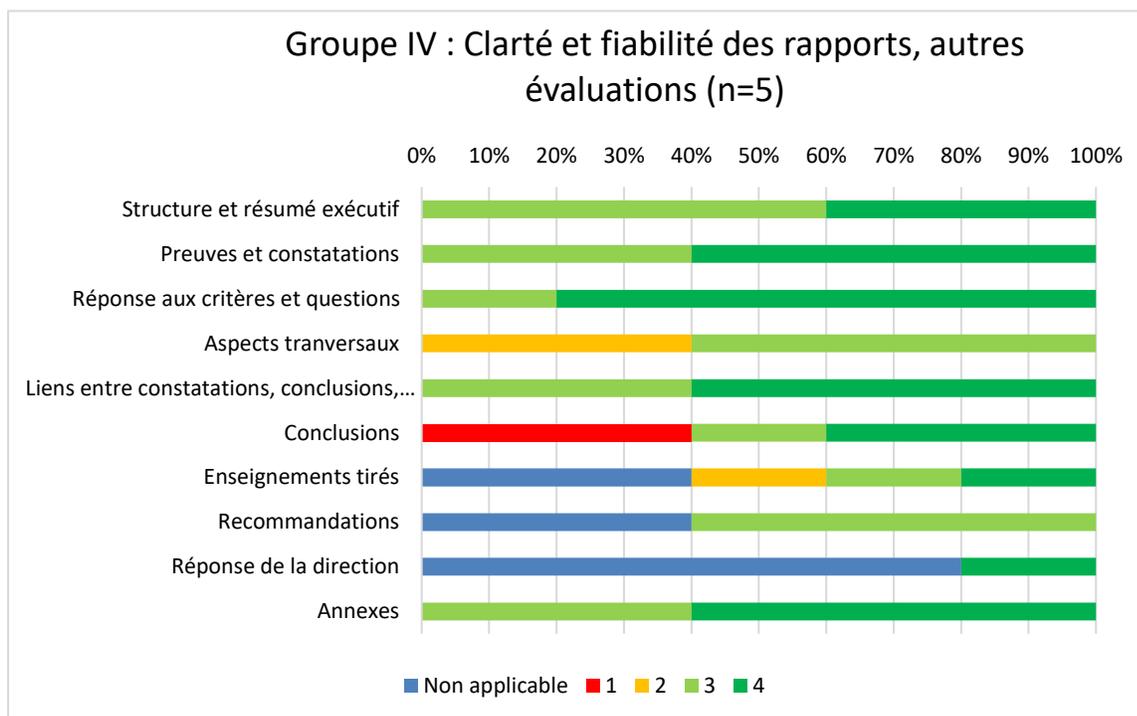
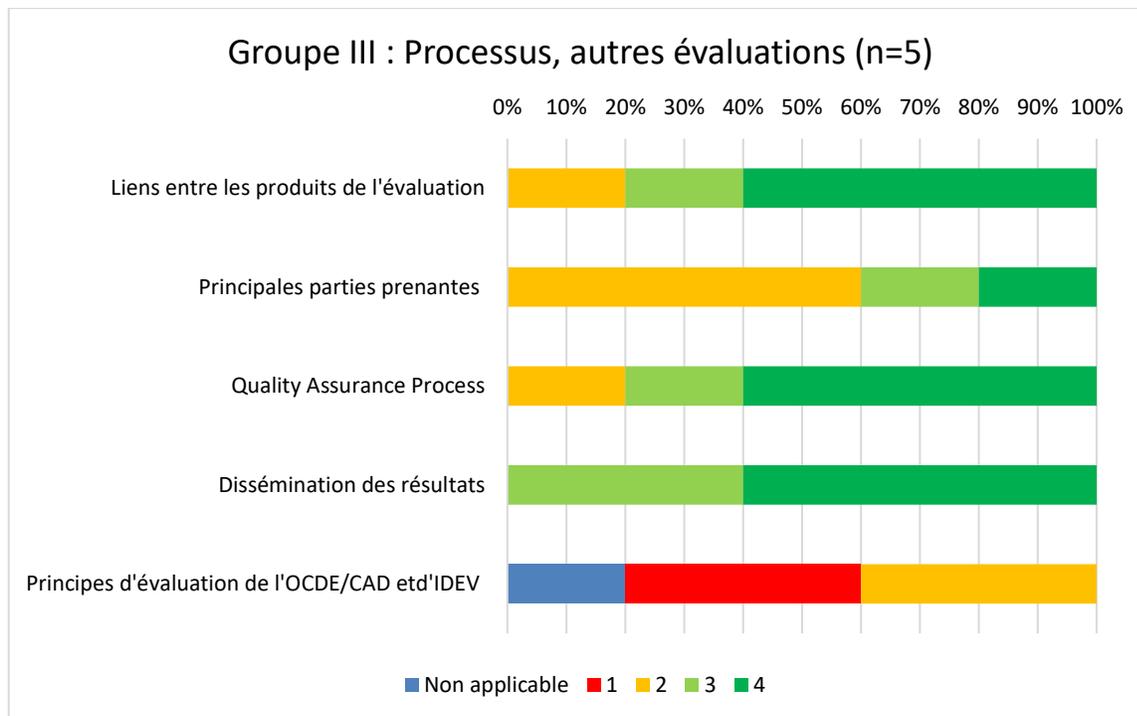
Groupe III : Processus, évaluations thématiques et sectorielles (n=4)



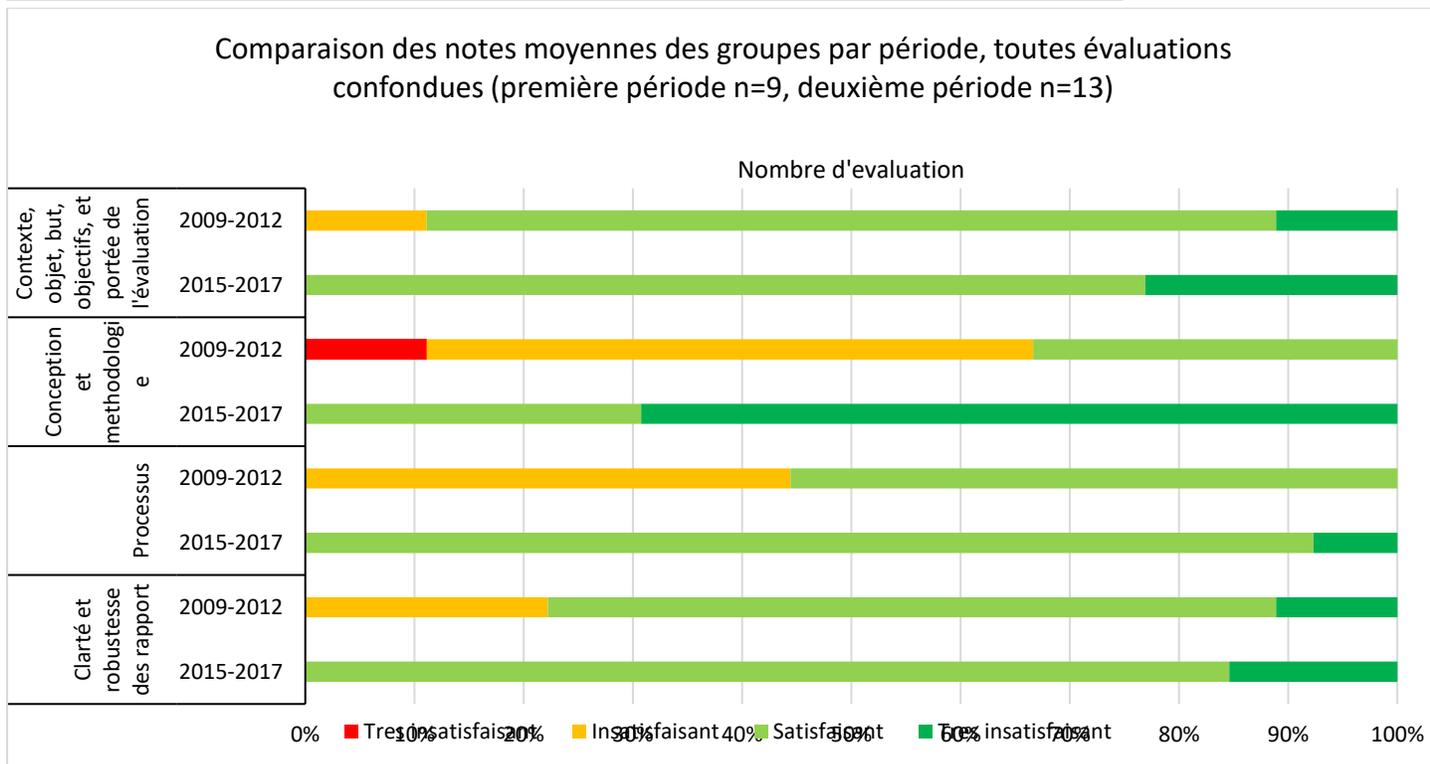
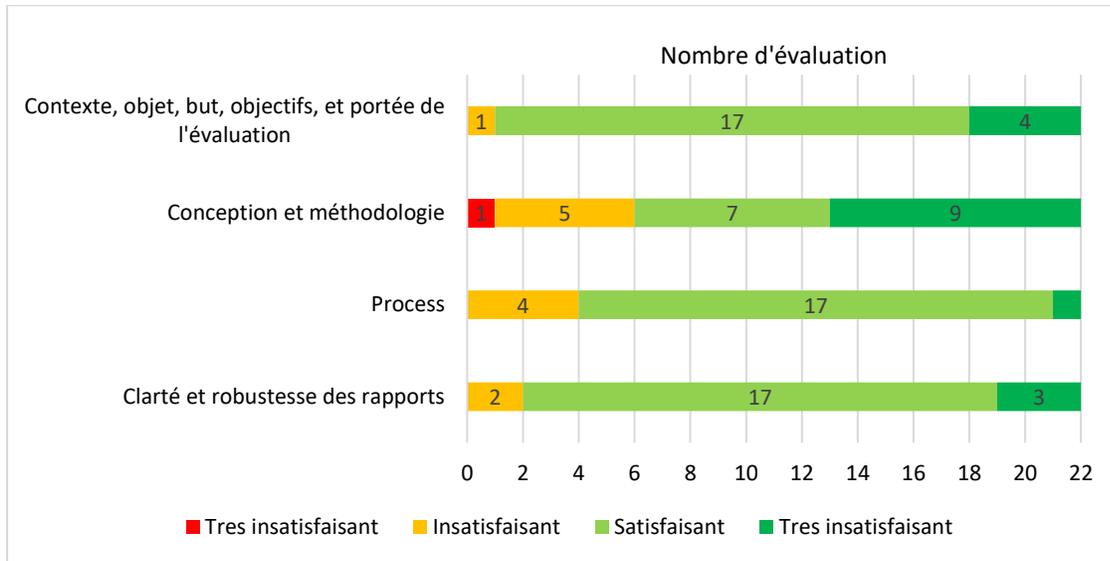


Annexe X. Résultats de l'appréciation de la qualité par critère, autres évaluations

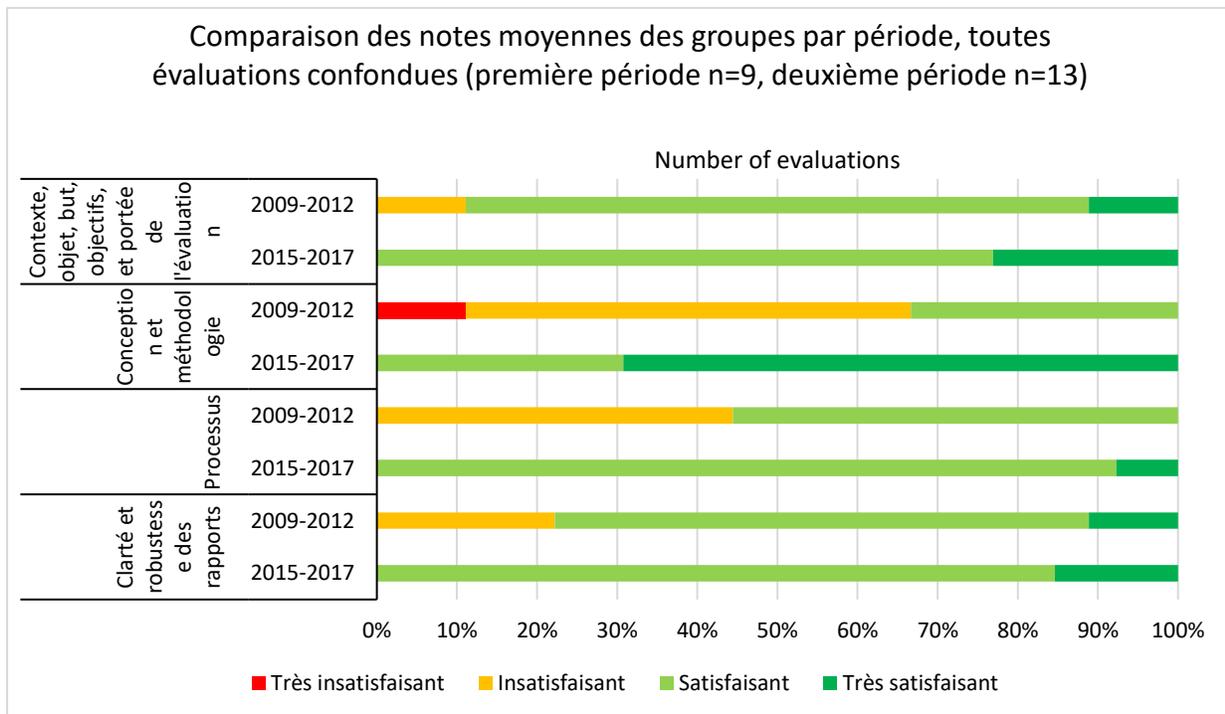
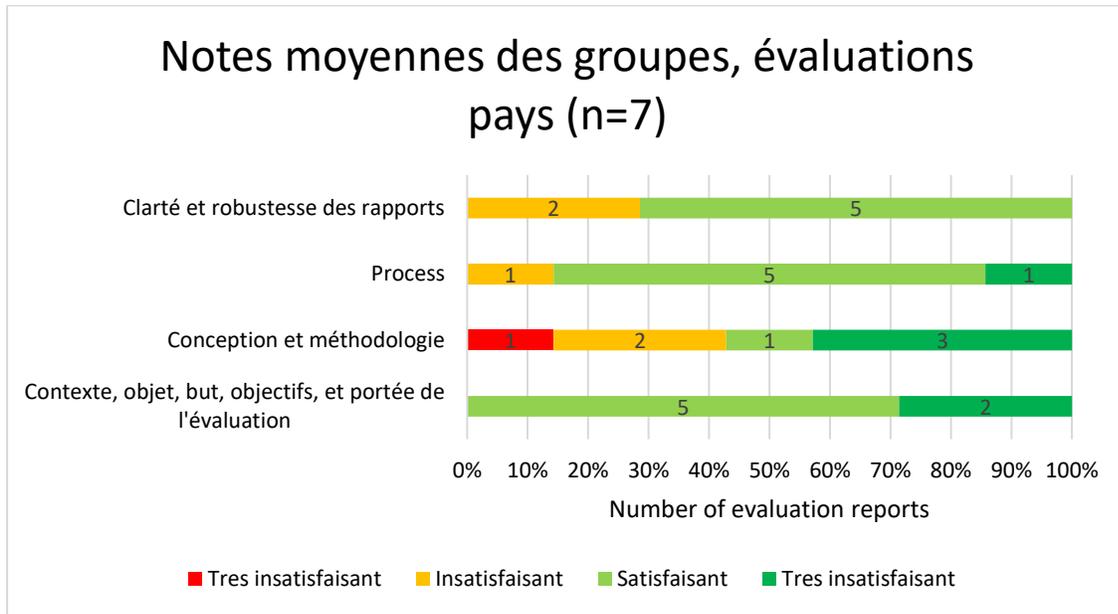




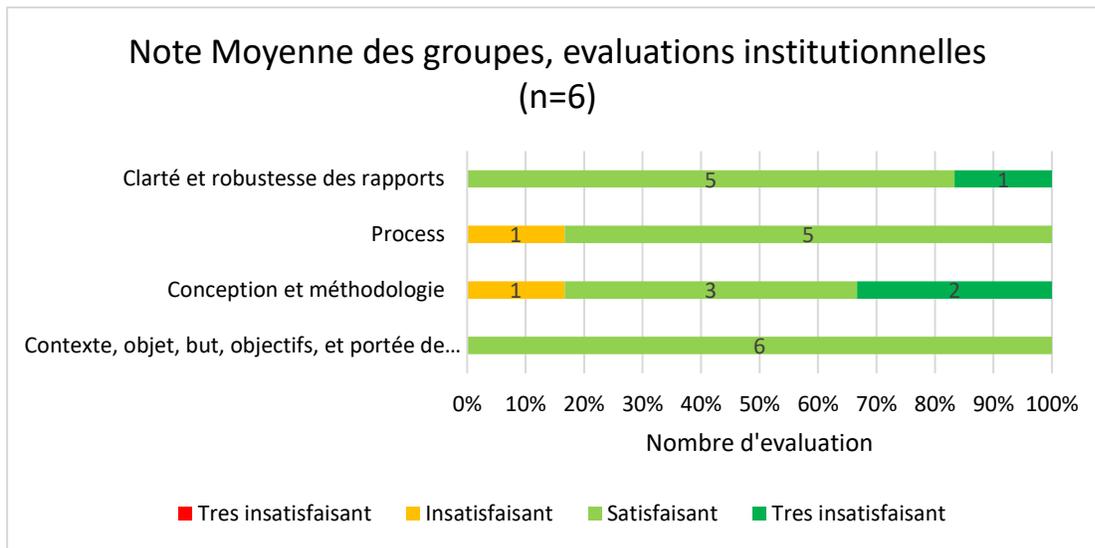
Annexe XI. Résultats de l'appréciation de la qualité par groupe de critères, toutes évaluations confondues

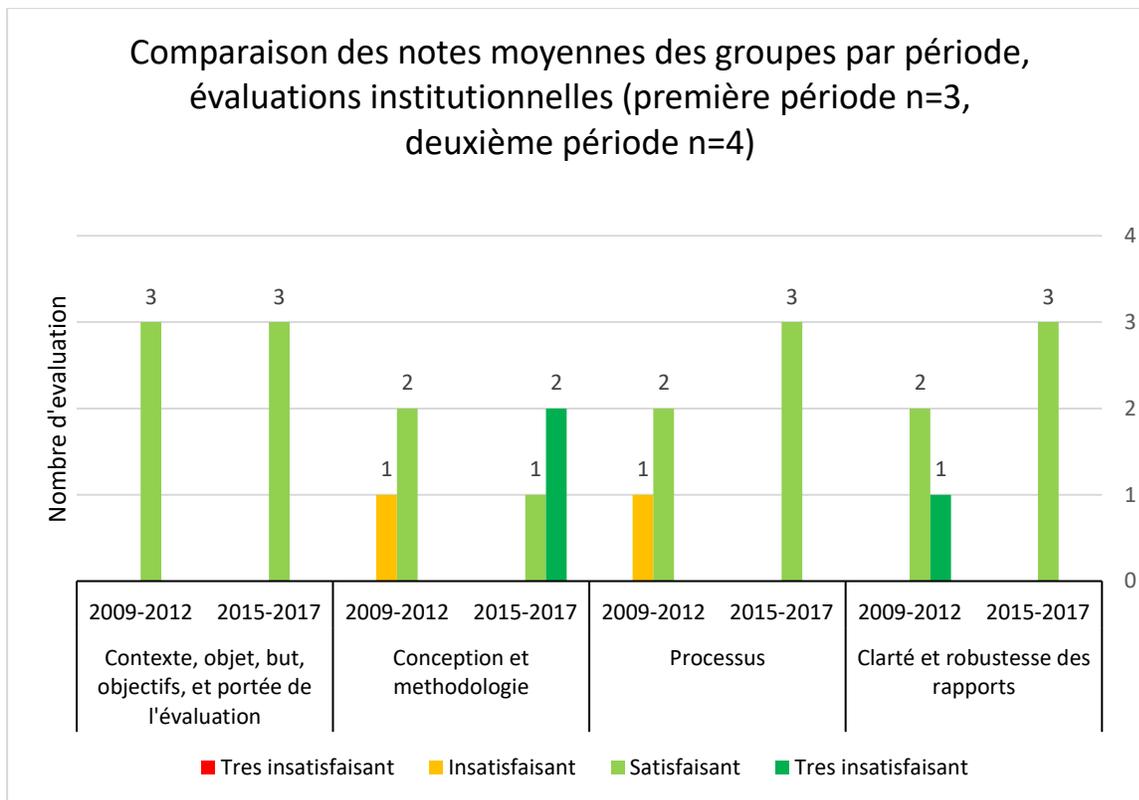


Annexe XII. Résultats de l'appréciation de la qualité par groupe de critères, évaluations par pays

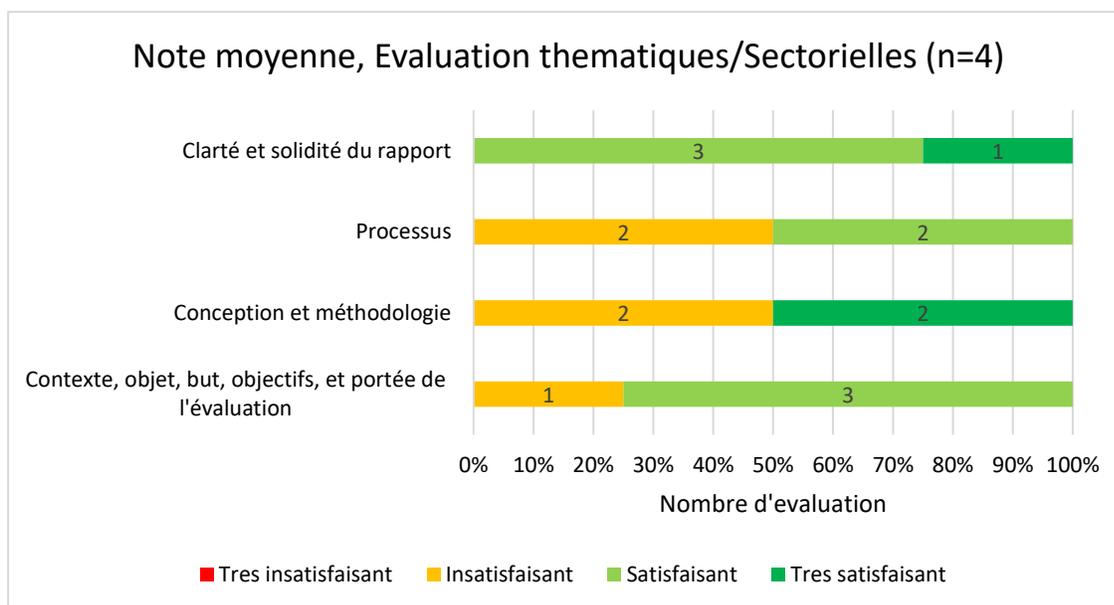


Annexe XIII. Résultats de l'appréciation de la qualité par groupe de critères, évaluations institutionnelles

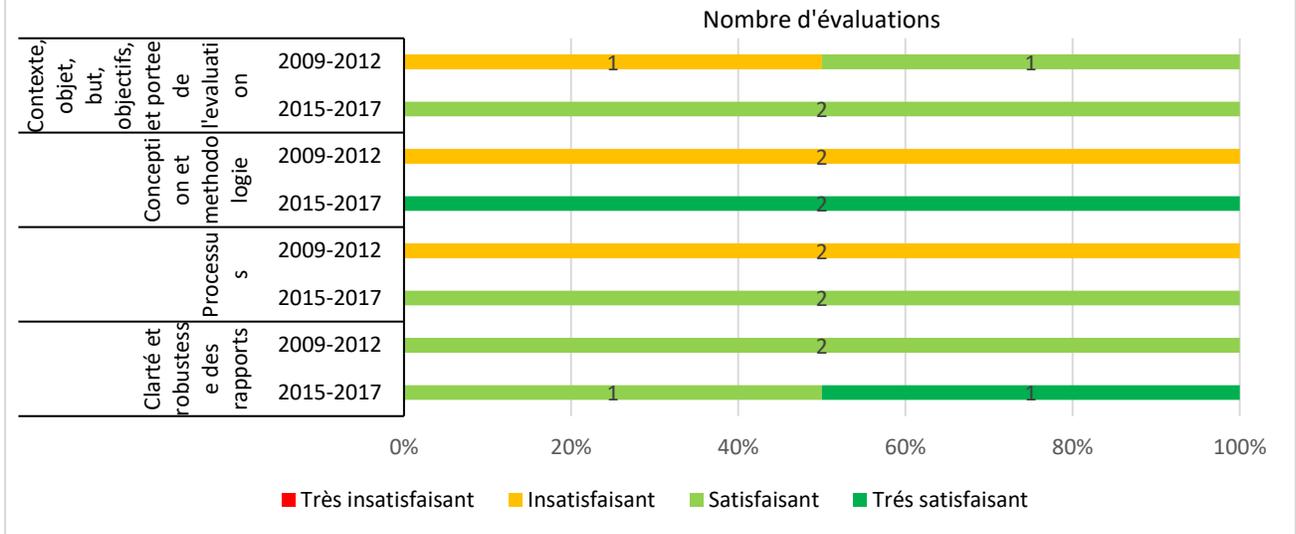




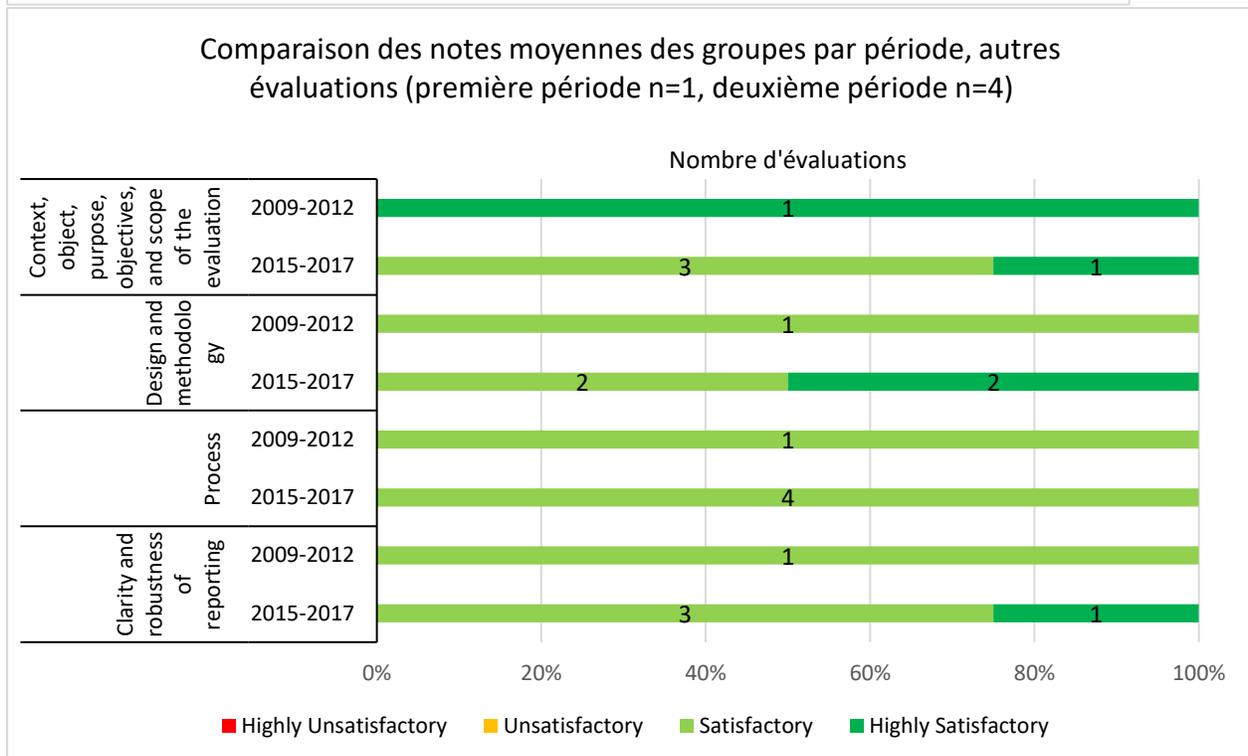
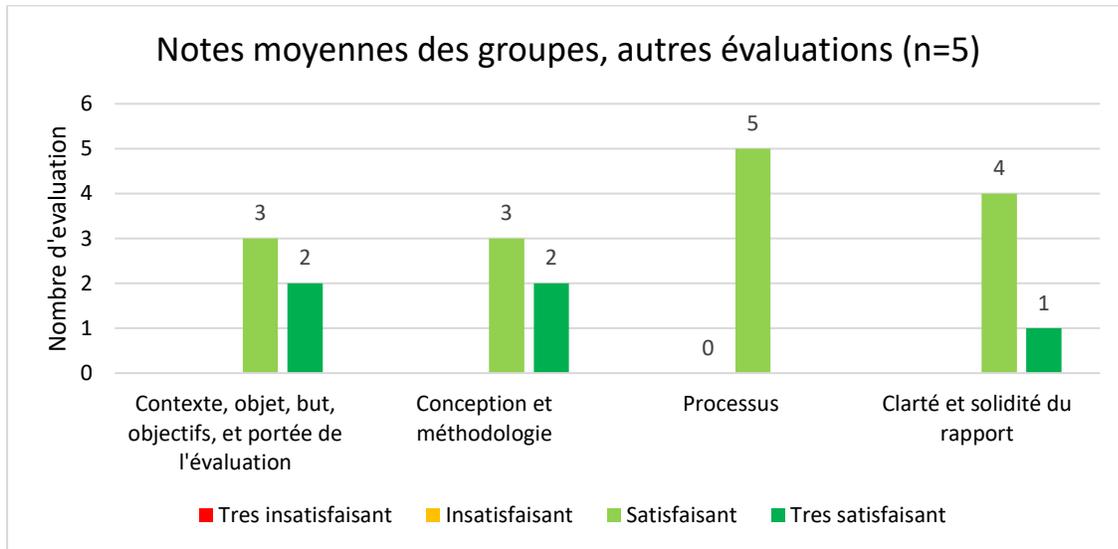
Annexe XIV. Résultats de l'appréciation de la qualité par groupe de critères, évaluations thématiques et sectorielles



Comparaison des notes moyennes des groupes par période, évaluations thématiques/sectorielles (première période n=2, deuxième période n=2)

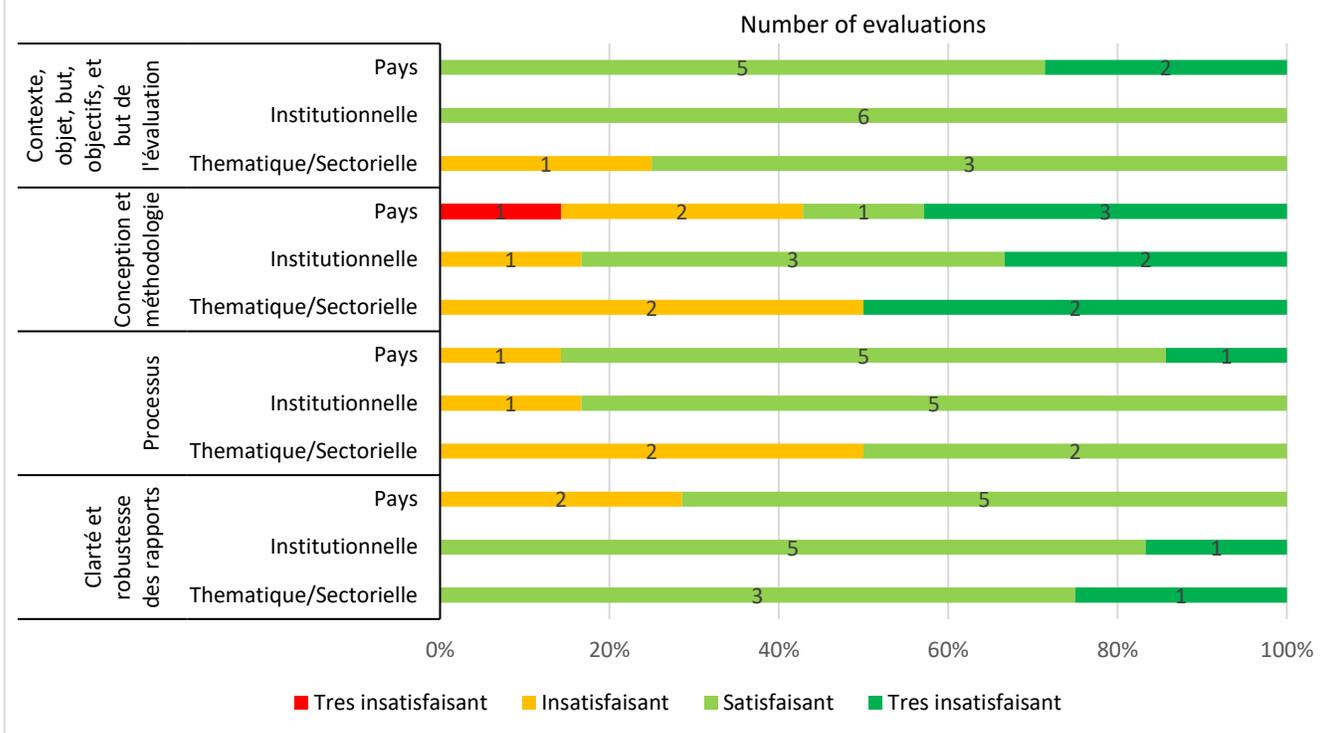


Annexe XV. Résultats de l'appréciation de la qualité par groupe de critères, autres évaluations



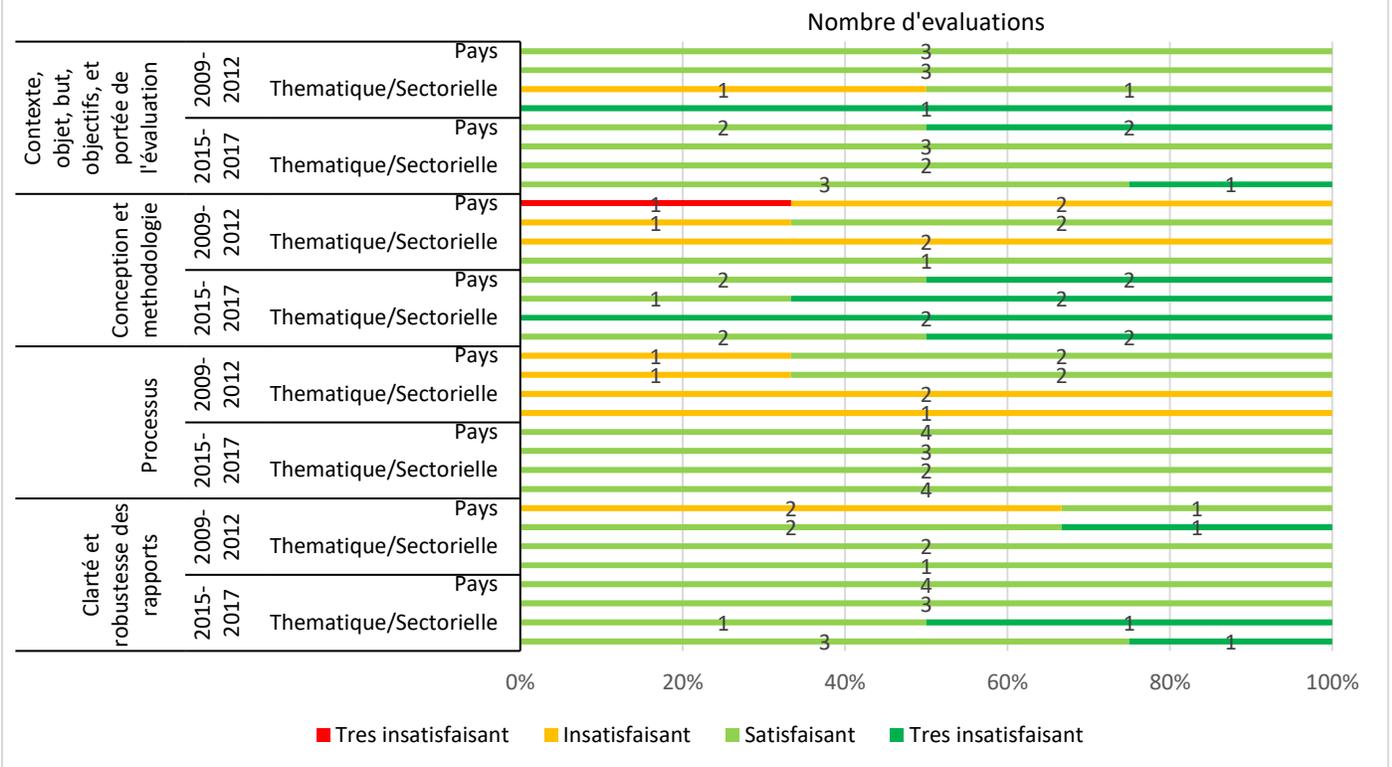
Annexe XVI. Résultats de l'appréciation de la qualité par groupe de critères et type d'évaluation

Comparaison des notes moyennes des groupes par type d'évaluation et période (pays n=7, institutionnelles n=6, thématiques/sectorielles n=4, autres n=5)



Annexe XVII. Résultats de l'appréciation de la qualité par groupe de critères, période d'évaluation et type d'évaluation

Comparaison des notes moyennes par groupes et par type d'évaluations et période (Paysn=7, Institutionnelle n=6, Thematique/Sectorielle n=4, Autres n=5)



Annexe XVIII. Liste des documents consultés

Normes internationales

Partenariat de Busan pour une coopération efficace au service du développement : Quatrième forum de haut niveau sur l'efficacité de l'aide, Busan, République de Corée, 29 novembre - 1er décembre 2011.

Groupe de coopération pour l'évaluation. 2017. *Integrating Gender into Project-level Evaluation (Intégrer le genre dans les évaluations de projets)*. Abidjan : BAD-IDEV

Groupe de coopération pour l'évaluation. 2012. *Big Book on Evaluation Good Practice Standards (Grand livre des normes de bonnes pratiques en matière d'évaluation)*. Manille

Groupe de coopération pour l'évaluation. 2008. *Normes de bonnes pratiques - Évaluations des stratégies et programmes par pays*. Manille

Institutionnelle

Groupe de la Banque africaine de développement. 2017. *Revue annuelle sur l'efficacité du développement 2017 Transformer l'Afrique — Libérer le potentiel de l'agriculture*. Abidjan

Banque africaine de développement. 2013. *Stratégie de la BAD pour la période 2013-2022 : Au centre de la transformation de l'Afrique* Abidjan

Banque africaine de développement - Fonds africain de développement. Juin 2001. *The Gender Policy (Politique en matière de genre)*. OESU.

Évaluation indépendante du développement 2013. *Stratégie d'évaluation indépendante de la Banque africaine de développement pour 2013-2017*. Groupe de la Banque africaine de développement. Abidjan

Évaluation indépendante du développement 2013. *Évaluation de la mise en œuvre de la Stratégie d'évaluation indépendante de la BAD 2013-17 - Rapport de démarrage V3*. Groupe de la Banque africaine de développement. Abidjan

Évaluation indépendante du développement 2016. *Évaluation globale des résultats de développement du Groupe de la Banque africaine de développement 2004-2013*. Rapport de synthèse. Groupe de la Banque africaine de développement. Abidjan

Évaluations de pays :

AFRIQUE DU SUD

- Réponse de la direction, Banque africaine de développement, janvier 2017.

- Évaluation de la stratégie et du programme par pays 2004-2015, Rapport de synthèse, IDEV, septembre 2016.
- Évaluation de la stratégie et du programme pays 2004-2015, Plan de diffusion préliminaire, 2016.
- Évaluation indépendante du programme et de la stratégie de la BAD en Afrique du Sud 2004-2014, Présentation, IDEV, 2016.
- Évaluation de la stratégie et du programme par pays 2004-2014, Rapport technique final, ECORYS - IDEV, mai 2016.
- Nedbank Limited - Ligne de crédit, Évaluation de la performance du projet, 2016.
- Nedbank Limited - Notes subordonnées de niveau II, Évaluation de la performance du projet, 2016.
- Standard Bank of South Africa, Évaluation de la performance du projet, 2016.
- Standard Bank of South Africa - Ligne de crédit, Évaluation de la performance du projet, 2016.
- Ligne de crédit I de la garantie non souveraine de l'Industrial Development Corporation, Évaluation de la performance du projet, 2016.
- Ligne de crédit II de la garantie non souveraine de l'Industrial Development Corporation, Évaluation de la performance du projet, 2016.
- Ligne de crédit de la banque de crédit agricole, Évaluation de la performance du projet, 2016.
- Banque de développement de l'Afrique australe, LdC IV, 2016.
- Transnet Limited, Évaluation de la performance du projet, 2016.
- Centrale électrique de Medupi, Évaluation de la performance du projet, 2016.
- Ferme éolienne de Sere, Évaluation de la performance du projet, 2016.
- Évaluation de la stratégie et du programme par pays 2004-2014, Rapport de démarrage - Final, ECORYS - IDEV, octobre 2015.
- Évaluation de la stratégie et du programme par pays 2004-2014, Demande de propositions, Banque africaine de développement, 2015.

CAMEROUN

- Évaluation de l'assistance de la Banque, Banque africaine de développement, Fonds africain de développement, novembre 2009.
- Évaluation de l'aide-pays, document d'orientation, Département de l'évaluation des opérations, 2009.

CAP VERT

- Évaluation de l'assistance de la Banque 1996-2007, Groupe de la Banque africaine de développement, Département de l'évaluation des opérations, octobre 2009.
- Note d'orientation, Évaluation de l'assistance de la Banque au Cap Vert, 2009.

- Évaluation de l'aide-pays, Département de l'évaluation des opérations, octobre 2009

CÔTE D'IVOIRE

- Réponse de la direction, Banque africaine de développement, janvier 2018.
- Évaluation de la stratégie et du programme de la Banque 2006-2016, Plan de communication et de diffusion, 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement (BAD) en Côte d'Ivoire 2006-2016, Rapport de synthèse, IDEV, septembre 2017.
- Évaluation de la stratégie et du programme de la Banque 2006-2016, Rapport de synthèse, IDEV, septembre 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement (BAD) en Côte d'Ivoire 2006-2016, Rapport technique révisé, ADE, juin 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement (BAD) en Côte d'Ivoire 2006-2016, Rapport technique Volume II - Annexes, ADE, avril 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement (BAD) en Côte d'Ivoire 2006-2016, Rapport de démarrage, Volume II - Annexes, janvier 2017.
- Projet de pont à péage Henri Konan Bédié, Évaluation des résultats du projet, 2017.
- Programme d'appui à l'amélioration de l'employabilité et de l'insertion des jeunes (PAAEIJ), Évaluation des résultats du projet, 2017.
- Projet d'appui institutionnel multisectoriel à la sortie de crise (PAIMSC) - Volet Agriculture, Évaluation des résultats du projet, 2017.
- Projet d'appui institutionnel multisectoriel à la sortie de crise (PAIMSC) - Volet éducation, Évaluation des résultats du projet, 2017.
- Projet d'appui institutionnel multisectoriel à la sortie de crise (PAIMSC) - Volet santé et VBG, Évaluation des résultats du projet, 2017.
- Programme d'appui au renforcement de l'inclusion et de la cohésion sociale (PARICS), Évaluation des résultats du projet, 2017.
- Programme de riposte contre Ebola (PRCE), Évaluation des résultats du projet, 2017.
- Projet de gestion intégrée du bassin versant du Gourou - phase d'urgence, Évaluation des résultats du projet, 2017.
- Programme d'urgence pour la restauration des services sociaux et administratifs de base (PURSSAB), Évaluation des résultats du projet, 2017.
- Évaluation des stratégies et programmes de la Banque Africaine de Développement (BAD) en Côte d'Ivoire, Termes de référence des services de consultation, juillet 2016.

GABON

- Réponse de la direction à l'évaluation de l'assistance au Gabon, Banque africaine de développement, avril 2011.
- Évaluation de l'aide-pays, 1996-2008, Département de l'évaluation des opérations, Groupe de la Banque africaine de développement, 2011.
- Document d'orientation, Évaluation de l'assistance de la Banque au Gabon, 2011.
- Évaluation de l'assistance de la Banque au Gabon, Termes de référence du consultant, 2011.
- Évaluation de l'assistance de la Banque, 1996-2008, Rapport provisoire, Groupe de la Banque africaine de développement, Département de l'évaluation des opérations, août 2009.
- Évaluation de la performance du multisecteur, 1996-2008, Rapport provisoire, Banque africaine de développement, Département d'évaluation des opérations, mars 2009.
- Évaluation de l'assistance de la Banque au Gabon, Termes de référence de la mission, Groupe de la Banque africaine de développement, décembre 2008.

GHANA

- Évaluation de la stratégie et du programme de la Banque 2002-2015, Rapport de synthèse, IDEV, mars 2017.
- Évaluation de la stratégie et du programme par pays 2002-2014, Rapport technique, BKP, révisé en juillet 2016.
- Projet de développement agricole des plaines d'Afram (PDAPA), Évaluation des résultats du projet, 2016.
- Projet de sensibilisation à la qualité et à la commercialisation des exportations (PSQCE), Évaluation des résultats du projet, 2016.
- Projet d'expansion de la centrale Takoradi II, Évaluation des résultats du projet, 2016.
- Programme d'appui à la réduction de la pauvreté et au climat des affaires (PARPCA), Évaluation des résultats du projet, 2016.
- Deuxième prêt d'appui à la réduction de la pauvreté (PARP II), Évaluation des résultats du projet, 2016.
- Troisième prêt d'appui à la réduction de la pauvreté (PARP III), Évaluation des résultats du projet, 2016.
- Projet multinational pour la création de zones indemnes de tsetse et de trypanosomiase en Afrique de l'Est et de l'Ouest - Volet Ghana, Évaluation des résultats du projet, 2016.
- Évaluation de la stratégie et du programme par pays 2004-2014, Rapport de démarrage, BKP, Projet, révisé en décembre 2015.

UGANDA

- Évaluation de l'aide-pays, 2001-2007, IEG/OPEV, mars 2009.

RÉPUBLIQUE DÉMOCRATIQUE DU CONGO

- Évaluation de la stratégie et du programme de la Banque 2004-2015, Rapport de synthèse, IDEV, mars 2017.
- Évaluation de la stratégie et du programme de la Banque 2004-2015, Rapport de Synthèse, IDEV, Mars 2017.
- Réponse de la direction, Banque africaine de développement, mars 2017.
- Évaluation des stratégies et programmes de la Banque africaine de développement (BAD) en République démocratique du Congo, 2004-2013, Rapport technique, ADE, juin 2016.
- Évaluation des stratégies et programmes de la Banque africaine de développement (BAD) en République démocratique du Congo, 2004-2013, Rapport de démarrage - version finale, ADE, octobre 2015.

Évaluations institutionnelles

Évaluation indépendante de la stratégie et du processus de décentralisation à la Banque africaine de développement, 2009

- Rapport de synthèse pour la revue à mi-parcours du FAD-11, Document d'information, OPEV, Fonds africain de développement, octobre 2009.
- Réponse de la direction, BAD/FAD, juillet 2009.
- Rapport du Conseil - Version finale, IOD/PARC, juin 2009.
- Rapport : Première version, mai 2009.
- Évaluation de la décentralisation, changement le plus significatif : toutes les histoires, OPEV, 2009.
- Rapport intermédiaire sur l'utilisation des histoires de changements les plus significatifs, IOD/PARC, janvier 2009.
- Note conceptuelle, OPEV, révisée en décembre 2008.
- Rapport de démarrage, décembre 2008.
- Termes de référence pour l'évaluation, 2008.

Revue indépendante de la qualité à l'entrée des opérations et des stratégies FAD 2005-2008, 2010

- Rapport de synthèse final, Groupe BAD/OPEV, mars 2010
- Réponse de la direction, BAD/FAD, mars 2010.
- Rapport technique final, ITAD, septembre 2009.
- Rapport de démarrage révisé, mai 2009.
- Termes de référence pour les services de conseil, Annexe II, 2009.
- Document d'orientation.

Supervision de projet à la BAD 2001-2008, 2010

- Note d'évaluation, BAD/OPEV, juillet 2010.
- Rapport technique, BAD/FAD, juin 2010.
- Réponse de la direction, BAD/FAD, mars 2010.
- Rapport final, BAD/OPEV, février 2010.
- Annexe technique I, Approche et méthodologie, juin 2009.
- Annexe technique III, Extraits des rapports bancaires et des manuels sur la supervision de projets, juin 2009.
- Document d'orientation, avril 2005.

Évaluation indépendante de l'élaboration et de l'application des politiques et stratégies, 2015

- Réponse de la direction révisée, BAD/FAD, février 2016.
- Rapport de synthèse, 2016.
- Évaluations des engagements, Plan de communication et de diffusion, 2015.
- Volume II, Études de cas, Centennial Group, avril 2015.
- Rapport de démarrage, septembre 2014.
- Document d'orientation, mars 2014.

Évaluation globale des résultats de développement du Groupe de la Banque africaine de développement 2004-2013, 2016

- Rapport de synthèse, IDEV, octobre 2016.

- Document d'information interne : Une analyse comparative qualitative de la théorie du changement de la banque pour l'EGRD : facteurs d'évaluation censés contribuer à la performance de la BAD au niveau des pays, IDEV, mai 2016.
- Rapport final, IDEV, avril 2016.
- Annexe B : Matrice d'évaluation, 2016.
- Note méthodologique de synthèse, IDEV, mars 2016.
- Théorie du changement, février 2016.
- Synthèse de l'EGRD, Modèle de pays, Note d'orientation, IDEV, septembre 2015.
- Synthèse de l'EGRD, Modèle de pays, IDEV, septembre 2015.
- Modèle d'évaluation de la performance du projet, 2015.
- Document d'orientation, OPEV, mars 2014.
- Note exploratoire, OPEV, octobre 2013.

Politique de gestion des ressources humaines et orientations stratégiques de la Banque africaine de développement : Une évaluation formative, 2017.

- Rapport de synthèse, Version expurgée, IDEV, novembre 2017.
- Plan de communication et de diffusion, 2017.
- Rapport technique, Final (V3), EY, juin 2017.
- Conclusions émergentes, Présentation, EY, 2017.
- Rapport de démarrage, version 4, EY, 2017.
- Termes de référence, 2017.

Évaluations thématiques/sectorielles

Évaluation des opérations d'appui aux réformes de la Banque africaine de développement, 1999-2009, 2010

- Réponse de la direction, BAD/FAD, avril 2011.
- Rapport final, OPEV/Groupe de la BAD, 2011.
- Étude de cas par pays : Éthiopie, Division des évaluations de haut niveau (OPEV. 2), mars 2011.
- Étude de cas par pays : Maroc, Division des évaluations de haut niveau (OPEV. 2), mars 2011.
- Étude de cas par pays : Rwanda, Division des évaluations de haut niveau (OPEV. 2), mars 2011.
- Étude de cas par pays : Sierra Leone, Division des évaluations de haut niveau (OPEV. 2), mars 2011.

- Étude de cas par pays : Tanzanie, Division des évaluations de haut niveau (OPEV. 2), mars 2011.
- Examen institutionnel et politique, Division des évaluations de haut niveau (OPEV. 2), décembre 2010.
- Analyse de documents et d'expérience, Division des évaluations de haut niveau (OPEV. 2), novembre 2010.
- Étude de cas : Prêt d'appui à la diversification économique du Botswana, Division des évaluations de haut niveau (OPEV.2), septembre 2010.
- Étude de cas : Programme de réforme du secteur financier en Égypte, Division des évaluations de haut niveau (OPEV.2), septembre 2010.
- Rapport de démarrage final, OPM/Mokoro/CSEA, mai 2010.
- Termes de référence, 2009.

Libération du potentiel des sous-régions de l'Afrique - Revue de l'assistance du Groupe de la Banque aux banques de développement sous-régionales, 2010.

- Rapport final, Division des projets et programmes (OPEV. 1), juillet 2010.
- Rapport de démarrage, OPEV, janvier 2009.

Évaluation de l'aide du Groupe de la Banque aux petites et moyennes entreprises (2006–2013), 2015.

- Réponse de la direction, BAD/FAD, septembre 2015.
- Rapport de synthèse, Version expurgée, IDEV, septembre 2015.
- Rapport technique d'évaluation, IDEV, décembre 2014.
- Projet de rapport de démarrage, octobre 2013.

Appui de la Banque africaine de développement au développement des chaînes de valeur agricoles : Enseignements pour la stratégie « Nourrir l'Afrique », 2017.

- Plan de communication et de diffusion, 2017.
- Rapport de synthèse, IDEV, décembre 2017.
- Rapport technique final, SDF Global, octobre 2017.
- Études de cas : rapport du Kenya (final), ADE, août 2017.
- Études de cas : rapport du Libéria (final), ADE, août 2017.

- Études de cas : rapport du Maroc (final), ADE, août 2017.
- Études de cas : rapport du Mozambique (final), ADE, août 2017.
- Études de cas : rapport du Rwanda (final), ADE, août 2017.
- Études de cas : rapport de l'Ouganda (final), ADE, août 2017.
- Études de cas : rapport de la Zambie (final), ADE, août 2017.
- Études de cas : rapport de la Côte d'Ivoire (final), ADE, août 2017.
- Projet de rapport de démarrage, IDEV, décembre 2016.
- Document d'orientation, 2016.

Autres : Synthèse

Intégration de l'égalité des genres : Une voie vers des résultats ou une voie vers nulle part ? 2012

- Synthèse de l'évaluation, Présentation, 2012.
- Rapport de synthèse, OPEV, 2012.
- Document d'orientation, OPEV, 2010.
- Mémoire inter-office, Groupe de la BAD, juin 2010.

Vers une croissance tirée par le secteur privé : Leçons de l'expérience - Rapport de synthèse d'évaluation, 2016.

- Rapport de synthèse d'évaluation, IDEV, octobre 2016.
- Projet de rapport de synthèse d'évaluation, IDEV, octobre 2016.
- Rapport de synthèse d'évaluation, Version révisée, Centennial Group, juillet 2016.
- Rapport de démarrage, Centennial Group, avril 2016.
- Termes de référence, 2016.

Autres : Impact

Évaluation d'impact du Programme d'approvisionnement en eau et d'assainissement en milieu rural en Éthiopie 2006–2014, 2016.

- Réponse de la direction, BAD/FAD, octobre 2016.
- Rapport de synthèse, IDEV, septembre 2016.
- Rapport, IDEV, septembre 2016.
- Rapport de démarrage, octobre 2014.
- Note conceptuelle d'évaluations d'impact d'OPEV sur l'assistance du Groupe de la BAD : 1. Éthiopie : approvisionnement en eau et assainissement en milieu rural ; 2. Secteur de l'énergie en Tanzanie, OPEV, mars 2014.
- Projet de document d'orientation, 2014.

Autres : Régionale

Évaluation de la stratégie et des opérations d'intégration régionale de la Banque africaine de développement en Afrique centrale, 2011-2016, 2017.

- Plan de diffusion des résultats, 2017.
- Rapport de synthèse, IDEV, décembre 2017.
- Étude de cas : Évaluation du programme d'appui à la conservation des écosystèmes du Bassin du Congo (PACEBCo), Version finale, novembre 2017.
- Étude de cas : Évaluation de l'appui au renforcement des capacités institutionnelles de la Communauté économique des États de l'Afrique centrale (PARCI-CEEAC), Version finale, novembre 2017.
- Étude de cas : Évaluation du programme de développement durable du bassin du lac Tchad (PRODEBALT), Version finale, novembre 2017.
- Analyse du portefeuille des opérations multinationales, novembre 2017.
- Étude de cas : Évaluation du projet de route Ketta-Djoug, Version finale, octobre 2017.
- Examen du document de stratégie d'intégration régionale (DSIR), Version finale, septembre 2017.
- Examen de la qualité à l'entrée, Version finale, juillet 2017.
- Défis et opportunités de l'intégration régionale en Afrique centrale, Version finale, décembre 2016.

- Concepts du régionalisme et leur utilisation dans les documents de politique et de stratégie de la Banque, décembre 2016.
- Document d'orientation, décembre 2016.

Autres : Groupée

Stimuler le développement socio-économique local à travers l'électrification rurale, 2017.

- Plan de communication et de diffusion, 2017.
- Rapport, IDEV, novembre 2017.
- Rapport de synthèse, IDEV, février 2016.
- Protocole commun de collecte de données pour l'évaluation de la performance de projet, février 2015.
- Document d'orientation sur les évaluations groupées, Projet, janvier 2015.

Annexe XIX. Programme et rapport de l'atelier d'Abidjan



Présentation de l'atelier :
Appréciation de la qualité d'un échantillon de produits d'évaluation
BAD-IDEV

Programme du 27 mars 2018

9h00-9h10	Présentation de l'atelier : M. Rakesh Nangia, Évaluateur général, IDEV
9h10-9h20	Présentation (Tous les participants)
9h20-9h30	Analyse des objectifs et des étapes de l'appréciation de la qualité (Penny et Anne-Marie)
9h30-9h40	Aperçu du programme et des objectifs de l'atelier (Anne-Marie)
9h40-10h30	Présentation PowerPoint interactive : principales constatations et recommandations de l'appréciation de la qualité (Anne-Marie et questions des participants) <ul style="list-style-type: none"> a) Forces et faiblesses des évaluations 2009-2013 comparées à celles de 2015-2017 par thème : <ul style="list-style-type: none"> • Contexte, objet, finalité, objectifs et portée de l'évaluation • Conception et méthodologie • Processus • Clarté et fiabilité des rapports b) Des points à clarifier ?
	<i>Pause (10h30-10h45)</i>
10h45-12h30	Discussion sur les recommandations de l'appréciation de la qualité, y compris les implications pour la gestion et la communication des connaissances (cinq sous-groupes thématiques) <ul style="list-style-type: none"> a) Est-ce que les recommandations d'AQ liées à chacun des quatre thèmes clés identifiés trouvent un écho auprès des participants ? b) Aller de l'avant - Comment les recommandations peuvent-elles : <ul style="list-style-type: none"> • Être mises en œuvre et soutenir l'apprentissage ? • Servir de bases à la prochaine stratégie quinquennale ?

- Aider IDEV à identifier les résultats à faible valeur ajoutée et les investissements à long terme ?

Déjeuner (12h15-14h00)

14h00-15h00	Rapport des cinq groupes et discussion avec toute l'équipe
15h00-15h30	Questions pour les participants en vue de la finalisation du rapport de synthèse (Anne-Marie)
15h30-16h00	Discussion autour des prochaines étapes pour qu'IDEV utilise ces résultats dans la pratique
16h00-16h15	Conclusion et évaluation de l'atelier

Résumé

L'atelier avec le personnel d'IDEV tenu au siège de la BAD à Abidjan le 27 mars 2018 était essentiellement un atelier interactif de partage des connaissances pour discuter des processus, des résultats et des recommandations de l'appréciation de la qualité menée par Universalia. Mme Anne-Marie Dawson a représenté l'équipe d'appréciation durant l'atelier, lequel était co-animé par Mme Penelope Jackson, responsable d'évaluation chez IDEV. Après une présentation de l'atelier par M. Rakesh Nangia, Évaluateur général chez IDEV, et la présentation de quelque 25 membres d'IDEV, le processus, la méthodologie, les principales constatations et les recommandations de l'appréciation de la qualité ont été présentés à l'aide d'un PowerPoint, qui a abouti sur une discussion des participants autour de diverses questions. Parmi celles-ci : la méthodologie de l'appréciation de la qualité, l'analyse des données, les mécanismes institutionnels pour communiquer les résultats de l'évaluation et donner suite aux recommandations, entre autres.

Après la présentation PowerPoint et la discussion, les participants se sont divisés en cinq sous-groupes, chacun étant chargé d'examiner une recommandation spécifique de l'appréciation de la qualité, d'échanger des points de vue sur la question traitée et d'explorer les moyens de la mettre en œuvre. Après la pause déjeuner, un rapporteur de chaque sous-groupe a présenté les résultats de la discussion de son groupe à l'ensemble de l'équipe. Cela a permis à tous les participants de se familiariser avec les pratiques existantes, par exemple les stratégies pour renforcer les capacités des évaluateurs locaux dans le cadre de divers types d'évaluations, et de discuter des moyens d'intégrer les bonnes pratiques dans les processus et les recommandations pour améliorer les pratiques. Dans l'ensemble, les participants ont accepté les recommandations et partagé des idées pour les affiner afin de maximiser leur utilité. Les questionnaires d'évaluation de l'atelier, anonymement remplis par presque tous les participants, soulignent leur volonté de donner leur avis sur les recommandations d'appréciation de la qualité, de se rencontrer en groupe pour partager des idées sur leur travail et apprendre les uns des autres.

Vous trouverez ci-après le rapport détaillé de l'atelier réalisé par IDEV, avec la liste des participants.

Atelier sur le partage des résultats de l'appréciation de la qualité d'un échantillon d'évaluations IDEV,

27 mars 2018.

Membres de l'équipe : Anne-Marie Dawson (Universalialia), Penelope Jackson et Latefa Camara (IDEV)

Rapport de l'atelier

1. Objectif de l'atelier

L'objectif principal de cet atelier était de partager avec l'équipe d'IDEV les résultats de l'exercice d'appréciation de la qualité d'un échantillon de produits d'évaluation d'IDEV. Celui-ci a été réalisé par une société externe (Universalialia) de consultants pour des raisons de crédibilité et d'impartialité. L'évaluation de l'appréciation de la qualité fait partie de l'évaluation plus large de la mise en œuvre par IDEV de la Stratégie d'évaluation indépendante quinquennale 2013-2017. L'atelier a donc cherché à obtenir les points de vue des collègues d'IDEV sur ce qui ressort de l'appréciation de la qualité, y compris les projets de recommandations.

2. Introduction et présentation

Après les remarques introductives de l'évaluateur général et la présentation de chaque participant, Anne-Marie, chef d'équipe d'Universalialia, a présenté les principaux résultats de l'exercice d'appréciation de la qualité. Quelques suggestions ont été faites par les participants concernant l'appréciation de la qualité, comme le remplacement des termes « initiale » et « finale » par d'autres expressions moins statistiques, comme « première période » et « deuxième période ». Le fait que la taille de l'échantillon ne permet pas de tests statistiques a également été noté.

La présentation a montré que, dans l'ensemble, Universalialia a constaté une amélioration de la qualité des évaluations produites par IDEV au cours des années, principalement en comparant les évaluations de la première période (2009-2012) et celles de la deuxième période (2015-2017).

Un participant a souligné le fait que beaucoup d'autres changements chez IDEV (à part pour ce qui est de la stratégie) peuvent avoir contribué à cette amélioration de la qualité, y compris l'augmentation du budget pour les évaluations au cours des dernières années. Il a été précisé que l'apport de tout changement à la stratégie ou autre ne faisait pas partie du mandat de la mission confiée à Universalialia. Les participants ont également souligné les restrictions institutionnelles, comme le nombre maximal de pages des rapports présentés au Conseil qui pourrait limiter la quantité d'informations incluses dans les « rapports de synthèse » par opposition à d'autres documents de référence.

3. Issue des discussions de groupe

Après la présentation et les questions/réponses, les participants de l'atelier ont été divisés en différents groupes pour mener une discussion plus approfondie basée sur les différentes recommandations faites par l'équipe d'Universalialia. Ces discussions en groupe ont été intéressantes grâce à l'engagement actif des participants. Sur les 8 recommandations formulées,

les groupes ont eu l'occasion d'en aborder cinq, en fonction de ce que les participants jugeaient plus pertinent et important pour le travail d'IDEV. Les principaux résultats ou questions abordés au cours de ce travail de groupe sont les suivants :

- ⇒ **Recommandation** : *Considérer l'ajout d'une cartographie des parties prenantes. Cette exigence devrait être stipulée dans les TdR et la cartographie des parties prenantes élaborée par les évaluateurs et présentée dans les rapports de démarrage et d'évaluation (en annexe si trop détaillée). De cette manière, toutes les parties prenantes essentielles seraient identifiées et leurs rôles et contributions clairement décrits.*

Conclusion du groupe : L'équipe est globalement d'accord avec cette recommandation et pense que cela renforcera la transparence et l'utilisation de l'évaluation. L'approche, pour ce qui relève de la cartographie des parties prenantes, pourrait être définie dans le manuel d'évaluation d'IDEV prévu dans le document d'orientation/rapport initial. Il est important d'identifier les parties prenantes en fonction des sources potentielles d'information pour l'évaluation, ainsi que de leurs différents besoins en matière d'information.

- ⇒ **Recommandation** : *Assurer une intégration transversale efficace des questions liées à l'égalité des sexes, à la croissance inclusive, à l'environnement/au changement climatique et à la fragilité, qui sont explicitement identifiées comme des domaines prioritaires dans la stratégie décennale de la BAD. Les constatations révèlent une couverture inégale de ces dimensions, à la fois pour chaque type de produit d'évaluation et pour l'ensemble des produits. Ces dimensions sont communément appelées « transversales », mais elles ne sont pas toujours efficacement intégrées de manière transversale dans les évaluations.*

Conclusion du groupe : IDEV devrait veiller à identifier les questions transversales qui importent pour le département et s'assurer que, pour chaque évaluation, la question transversale pertinente est identifiée et prise en compte. Il est important de noter que toutes les questions transversales ne sont pas pertinentes pour chaque type d'évaluation. L'intégration des questions transversales dépend aussi de la conception des opérations. Parfois, il n'y a pas de données, même quand l'équipe d'évaluation veut faire avancer ces questions, car la conception ou la mise en œuvre/le suivi n'en a pas tenu compte. Cependant, des occasions d'évaluer certaines questions transversales existent, et nous devrions examiner cela au cas par cas.

- ⇒ **Recommandation** : *Veiller à ce que les principes d'évaluation de l'OCDE-CAD et de la BAD-IDEV, relatifs à l'indépendance, à l'impartialité, à l'éthique, au partenariat, à la coordination et au renforcement des capacités soient clairement et explicitement pris en compte dans le processus d'évaluation. Les documents d'orientation et rapports de démarrage devraient inclure une section standard concernant l'examen de ces principes.*

Conclusion du groupe : Cette disposition a été considérée comme une recommandation importante par le groupe, bien qu'il ait été reconnu qu'il y avait une différence entre ce qu'il y a sur le papier et ce qui existe en réalité (par exemple, l'indépendance et l'impartialité sont déjà énoncées dans la politique).

Les évaluations d'IDEV devraient par exemple s'assurer que le renforcement des capacités devienne une composante à part entière, en impliquant des consultants locaux dans les évaluations. C'est un point intéressant mais certaines questions ont été soulevées à ce sujet. Par exemple, dans la pratique, la mise en œuvre d'une telle initiative pour renforcer les capacités peut prendre beaucoup de temps lors d'une évaluation. La meilleure façon d'y parvenir est de bien planifier le processus et de l'inclure dans la conception de l'évaluation, afin que le temps alloué à l'ensemble de l'évaluation soit suffisant pour permettre le renforcement des capacités. C'est un processus très lourd tant administrativement qu'en termes de temps. Cependant, parfois, le renforcement des capacités qu'IDEV effectue par exemple au niveau des ministères est une chose différente. Il n'est pas toujours facile d'impliquer ces personnes dans notre évaluation en raison de la nature indépendante de ce que nous faisons.

IDEV devrait appliquer certaines pratiques de recherche à son évaluation. Par exemple, les questions d'éthique devraient être incluses et mentionnées lors de nos sondages ou de nos groupes de discussion/entrevues.

⇒ **Recommandation :** *Pour limiter le chevauchement entre les produits successifs pour chaque évaluation, envisager de remplacer les documents d'orientation par des termes de référence moins prescriptifs et exiger que les rapports de démarrage soient systématiquement produits par des consultants externes détaillant la méthodologie d'évaluation et le plan de travail.*

Conclusion du groupe :

Il est important qu'IDEV distingue les différents types de documents que l'unité doit produire au cours d'une évaluation et précise ce que chaque document doit contenir. Un manuel IDEV, détaillant chaque produit que nous livrons et dans quelles circonstances ils sont requis, est nécessaire. La nécessité de rédiger un rapport de démarrage et un document d'orientation au lieu d'un rapport de démarrage seul dépend par exemple de chaque évaluation. Par exemple, en présence d'un document d'orientation suffisamment détaillé, une équipe pourrait se dire qu'il est inutile de rédiger en plus un rapport de démarrage ; ou, si elle fait appel à une équipe de consultants individuels, elle pourrait demander à ces consultants de rédiger des rapports de démarrage pour leurs sujets spécifiques. Certains pensent que le document d'orientation devrait être rédigé par l'équipe d'IDEV, et que le rapport de démarrage devrait ajouter plus de valeur en venant compléter ce qui est mis à disposition des consultants externes. Certains estiment que remplir les rapports de démarrage en interne laissait plus de possibilités, d'autres considèrent qu'utiliser une entreprise qui n'avait pas intégré la méthodologie en participant à la phase de démarrage représentait un risque.

Ces différents points de vue montrent bien qu'IDEV a plusieurs façons d'impliquer des consultants. Il est possible de sous-traiter toute la collecte de données comme il est possible de

ne sous-traiter que des composants spécifiques ; il est aussi envisageable d'utiliser des entreprises ou des individus. Les documents produits doivent se rapporter à ce qui fonctionne pour chaque modèle. Cela doit être clairement documenté.

⇒ **Recommandation** : *Bien distinguer les constatations, les conclusions et les leçons apprises de l'évaluation. Les constatations relatives aux rapports d'évaluation examinés font parfois état d'un chevauchement entre ces éléments, par exemple, des constatations incluant des jugements d'évaluateurs et des notes de performance, des conclusions reprenant essentiellement les constatations clés, ou encore des leçons apprises formulées comme des conclusions, voire même des recommandations.*

Conclusion du groupe : Il est important que les évaluateurs d'IDEV aient une définition claire de ce que chaque concept signifie. Cela garantira une distinction claire et logique entre constatations, conclusions et recommandations. Les recommandations devraient être liées aux conclusions. Les leçons apprises sont importantes, mais les quatre aspects (constatations, conclusions, leçons et recommandations) doivent-ils faire partie d'une évaluation ?

4. Marche à suivre

Sur la base des différentes discussions, l'équipe d'Universalialia examinera certains aspects du rapport d'appréciation de la qualité avant de soumettre une nouvelle version à IDEV. Les questions abordées aujourd'hui seront prises en compte à la fois dans les appréciations et dans les recommandations. Ce document sera partagé avec toute l'équipe. L'appréciation des évaluations individuelles sera partagée avec les chargés de gestion. La version finale du rapport d'appréciation de la qualité servira à l'auto-évaluation globale de la mise en œuvre de la stratégie IDEV.

Annexe 1 : Aperçu des commentaires des participants à l'atelier

Dans l'ensemble, l'atelier a été bien accueilli par les participants. Plus de 80 % des commentaires étaient d'accord ou fortement d'accord sur le fait que les objectifs de l'atelier étaient clairs et pertinents. Les participants ont également convenu qu'ils avaient eu suffisamment de temps et d'espace pour exprimer leurs opinions.

Parties les plus intéressantes de l'atelier :

- Les discussions de groupe sont une des composantes les plus intéressantes de l'atelier, puisqu'elles ont permis aux participants d'exprimer des points de vue sur des points d'intérêt spécifiques. Les discussions sur les recommandations ont été jugées particulièrement intéressantes par les participants. L'opportunité de discuter des recommandations issues de l'appréciation de la qualité a été très appréciée.

Principaux points à retenir de l'atelier

Points à retenir les plus intéressants :

- L'importance de la stratégie dans l'amélioration des résultats
- L'importance du travail d'équipe
- L'intégration de thèmes transversaux dans les évaluations
- Discussions sur les défis rencontrés par les évaluateurs
- Rappel de la nécessité de maintenir un certain niveau de rigueur dans nos évaluations

Annexe 2 : Liste des participants

N.	Nom	Fonction
1	Akua Arthur-Kissi	Chargée d'évaluation - IDEV2
2	Anne-Marie Dawson	Chef d'équipe Universalia
3	Brenda Tsausagae	Consultante - IDEV1
4	Carla Felix	Consultante - IDEV2
5	Clément Banse	Chargé d'évaluation - IDEV1
6	Daniel Andoh	Consultant - IDEV3
7	Eneas Gakusi	Chargé d'évaluation en chef - IDEV1
8	Éric Yai	Consultant - IDEV2
9	Eustace Uzor	Consultant - IDEV2
10	Foday Turay	Chargé d'évaluation en chef - IDEV1
11	Gilbert Adjimoti	Consultant - IDEV2
12	Girma Kumbi	Chargé d'évaluation principal - IDEV2
13	Innocent Bledou	Consultant - IDEV2
14	Joseph Mouanda	Chargé d'évaluation principal - IDEV1
15	Karen Rot-Munstermann	Chef de division - IDEV3
16	Latefa Camara	Consultante - IDEV2
17	Mabarakissa Diomande	Chargée d'évaluation -

		IDEV1
18	Madhu Mampuzhasseril	Chargé d'évaluation en chef - OIC IDEV2
19	Mireille Cobinah	Assistante aux évaluations (gestion des connaissances) - IDEV3
20	Mohamed Coulibaly	Consultant - IDEV1
21	Monica Lomena-Gelis	Chargée d'évaluation principale - IDEV2
22	Nagnouma Kone	Consultante - IDEV3
23	Najade Lindsay	Consultante - IDEV3
24	Penelope Jackson	Chargée d'évaluation en chef - IDEV1
25	Rakesh Nangia	Évaluateur général
26	Souleymane Dieye	Assistant supérieur de base de données - IDEV3
27	Stéphanie Yoboue	Consultante - IDEV1
28	Telesphore Some	Consultant - IDEV3