

# Entretiens

## Entretien avec Jean-Paul Moatti et Rémy Rioux

*Rémy Rioux, vous dirigez une institution (l'Agence française de développement – l'AFD) dont la mission est de financer des projets, des programmes et des politiques de développement. En quoi consiste l'évaluation dans votre institution et quel rôle joue-t-elle dans votre activité et celle de vos partenaires ?*

Pour bien comprendre le rôle de l'évaluation, il convient d'abord de la situer dans le cycle de travail de l'Agence française de développement à la lumière de ce que j'appelle, dans mon livre *Réconciliations*, l'approche des 4E pour *Écoute*, *Expertise*, *Engagement* et *Évaluation*. Le point de départ de l'identification par l'AFD d'un projet ou d'une politique publique est toujours l'expression d'un besoin par un partenaire d'un pays du Sud, comme un ministère, une collectivité locale ou un organisme de coopération internationale. Si ce besoin est en phase avec les priorités d'aide au développement qui ont été fixées à notre agence par le gouvernement français, nous débutons alors une phase d'écoute et de dialogue dans laquelle l'AFD s'appuie sur des expertises internes et externes pour analyser l'opportunité du projet, ainsi que sa faisabilité sur les plans technique, financier et institutionnel. Cette phase de conception est essentielle, car elle conditionne le champ de l'évaluation du projet en définissant, avec les parties prenantes, les objectifs de transformation poursuivis par le projet.

Le projet est ensuite passé au crible d'un « avis développement durable », établi par une structure indépendante des opérations, qui note le projet selon une grille d'analyse composite afin d'estimer les impacts potentiels sur six dimensions du développement durable. L'évaluation ne porte pas seulement sur les réalisations matérielles, mais aussi – et surtout – sur l'atteinte de l'objectif du projet. Nos actions de développement doivent en effet être évaluées par rapport à ce qu'elles apportent au tissu économique, environnemental et social local. Par exemple, si un projet vise à augmenter le taux de scolarisation des enfants et à améliorer leurs acquis éducatifs, la construction d'établissements fonctionnels et pérennes est une condition importante, mais non suffisante pour le succès

du projet. Le projet ne sera considéré comme réussi que si ces bâtiments sont fortement fréquentés et que l'on constate une amélioration significative des compétences des élèves en fin de cursus scolaire. Par conséquent, l'évaluation n'est possible que si la logique du changement a été clairement définie dès la conception du projet et si des critères de réussite, qui se traduisent par des indicateurs quantitatifs et qualitatifs renseignés de manière méthodique, ont été explicitement identifiés dès le départ.

Enfin, un dispositif de suivi doit être mis en place avant de démarrer la mise en œuvre du projet afin de contrôler non seulement le bon déroulement des activités financées, mais aussi l'atteinte des résultats intermédiaires. Si l'on reprend l'exemple de notre projet éducatif, cela signifie que l'on doit s'assurer, avant qu'il ne touche à sa fin, que les enseignants ont été formés et affectés dans les établissements construits dans le cadre du projet, et que les inégalités d'accès entre les filles et les garçons se réduisent. Si ce n'est pas le cas, il faudra prendre des mesures correctives avant l'achèvement du projet pour atteindre les objectifs fixés.

L'évaluation joue donc un rôle éminemment opérationnel. C'est un élément clé du métier de développeur pour comprendre les paramètres qui ont permis, accéléré ou freiné le déroulement du projet et la réalisation de ses objectifs, en tirer des enseignements et être plus efficace. Mais l'évaluation joue également un rôle de plus en plus stratégique. À l'heure où les besoins d'investissements pour réduire les inégalités et préserver les équilibres de notre planète vont croissant, la question principale qu'un bailleur de fonds comme l'AFD se pose est la suivante : comment l'évaluation peut-elle nous servir à coconstruire des projets de développement avec nos partenaires, et à soutenir des politiques publiques plus efficaces et efficientes au bénéfice des populations ? L'une des leçons que je tire de notre expérience à l'AFD est qu'évaluer notre action, c'est s'inscrire dans la démarche de redevabilité et d'amélioration continue. Voilà pourquoi l'évaluation joue un rôle central et croissant dans notre institution.

Il faut reconnaître qu'en raison de son histoire, l'évaluation de l'aide française a longtemps été reléguée au second plan par rapport aux autres pays et aux bailleurs multilatéraux. Les études comparatives de LAPORTE (2015) montrent que les pratiques d'évaluation ont émergé parallèlement en France et dans les pays anglo-saxons dans les années 1960, avec une approche spécifique en France, marquée par le rôle de l'Institut national de la statistique et des études économiques (Insee) et des experts en agronomie et développement rural de la coopération française. La France a toutefois pris du retard dans les années 1980 et 1990, à une époque où les bailleurs anglo-saxons systématisaient l'évaluation à mesure qu'augmentaient leurs budgets consacrés à l'aide internationale.

La France rattrape maintenant ce retard sous l'effet conjoint d'investissements croissants consacrés au développement durable et des exigences de résultats qui y sont associées. Les moyens dont dispose l'AFD augmentent progressivement pour atteindre l'objectif fixé par le Président français de consacrer 0,55 % du PIB à l'aide publique au développement d'ici 2022. Ces moyens accrus

s'accompagnent naturellement d'un impératif : celui de rendre toujours mieux compte, au gouvernement, aux parlementaires et aux citoyens, de l'efficacité de notre action. L'évaluation joue un rôle déterminant pour répondre à cette exigence de redevabilité.

*Parmi ses évaluations, l'AFD mène des études d'impact, qui ont une ambition scientifique plus poussée. Quelle est l'utilité de ces recherches évaluatives, et que pensez-vous de l'évaluation par assignation aléatoire (Randomized Controlled Trial – RCT), qui est présentée aujourd'hui comme la méthode la plus rigoureuse en la matière ?*

Dans ma réponse précédente, je me référais aux évaluations en général. Les études d'impact constituent un type particulier d'évaluation, qui est conçu pour identifier, mesurer et comprendre de manière scientifiquement rigoureuse les effets strictement imputables à une intervention. Elles s'appuient pour cela sur un contrefactuel, c'est-à-dire qu'elles mettent en regard l'évolution d'une population bénéficiaire à celle d'une population non bénéficiaire, en s'assurant que ces deux populations sont effectivement comparables et que l'intervention évaluée est le seul critère qui les différencie.

Les évaluations d'impact sont essentiellement utilisées comme un outil ponctuel pour démontrer (*to prove*), plutôt que comme un outil d'amélioration (*to improve*). Elles permettent d'établir qu'en général certains types d'interventions fonctionnent dans des contextes donnés, mais elles s'avèrent trop lourdes, longues et coûteuses pour être systématisées comme un outil de redevabilité et d'apprentissage. Pour ceux-ci, on dispose de méthodes plus légères et harmonisées, qui ont été régulièrement améliorées au cours des trois dernières décennies. Ces démarches peuvent s'appuyer sur de l'expertise locale existante et être mobilisées de manière plus agile pour alimenter le dialogue avec nos partenaires.

L'AFD réalise des études d'impact depuis le début des années 2000 pour participer à l'enrichissement d'un corpus de connaissances générales sur le développement. L'analyse comparative détaillée que nous avons élaborée sur ce sujet (PAMIÈS-SUMNER, 2015) montrait qu'il s'agissait d'une particularité de notre institution. À l'exception du Department for International Development (DFID), la plupart des bailleurs bilatéraux restent en retrait en matière d'évaluations d'impact par rapport aux grands bailleurs multilatéraux qui en ont produit des centaines.

Depuis quinze ans, l'AFD a également contribué activement à la réflexion méthodologique de la communauté du développement sur les évaluations d'impact. Parmi les méthodes qui peuvent être utilisées pour renforcer la comparabilité entre population bénéficiaire et contrefactuel pour déduire l'impact d'une intervention, la méthode des RCT a connu un fort engouement sur lequel le présent ouvrage se propose de prendre du recul. L'AFD a joué un rôle clé dans l'amorce de cette tendance en soutenant, dès 2005, deux vastes RCT sur des secteurs d'intervention qui étaient alors au cœur de l'actualité de l'aide : le microcrédit

et l'assurance santé. Ces études nous ont permis de faire progresser l'état des connaissances sur ces deux secteurs d'intervention, mais aussi sur les méthodes expérimentales, dont nous avons alors tiré un bilan mitigé (BERNARD *et al.*, 2012). Avec ces travaux, nous avons participé à l'émergence d'un consensus sur la nécessité d'évoluer vers des approches pluridisciplinaires combinant méthodes quantitatives et qualitatives.

Aujourd'hui, il est plus indispensable que jamais de mener des actions efficaces. Il ne reste que neuf ans pour atteindre les Objectifs de développement durable (ODD). Les interventions publiques doivent s'appuyer sur des connaissances empiriques si l'on veut répondre aux exigences légitimes des citoyens, dans un monde déchiré par des fractures sociales, politiques, économiques et, bien sûr, environnementales. En tant que plateforme pour le développement durable, l'AFD se doit d'être pragmatique et efficace, et nous sommes déterminés à réaliser davantage d'évaluations d'impact. Nous avons les moyens, objectifs et scientifiques, de savoir ce qui fonctionne et ce qui ne fonctionne pas.

Il reste une marge de progression. Comme je le souligne dans *Réconciliations*, les approches actuelles d'évaluation de l'impact de l'aide au développement sont loin d'être parfaites. Les résultats sont généralement mesurés par un instantané, utile, mais insuffisant, d'indicateurs quantitatifs tels que le nombre de personnes raccordées au réseau électrique, le taux de jeunes filles scolarisées ou les tonnes de CO<sub>2</sub> économisées. Les approches quantitatives sont souvent décevantes, car elles ne permettent pas d'identifier dans quelle mesure un projet de développement contribue aux dimensions qualitatives de l'agenda 2030. À cet égard, il est essentiel de définir un cadre commun pour déterminer les investissements qui sont en phase avec les trajectoires durables à long terme et ceux qui ne le sont pas, afin d'améliorer les approches qualitatives de l'évaluation pour atteindre les objectifs de l'accord de Paris et des ODD.

Comme les méthodologies se sont diversifiées, l'important est maintenant d'utiliser les approches les plus pertinentes pour répondre aux questions sur les sujets étudiés, sans préjugés théoriques. C'est la raison pour laquelle les équipes de l'AFD ont le souci constant de combiner les études d'impact et les évaluations opérationnelles plus légères en développant un ensemble diversifié d'outils de mesure permettant de mieux appréhender l'efficacité de l'aide. Nous prévoyons en particulier de mettre davantage la science au service de l'évaluation, en collaborant avec des partenaires comme l'Institut de recherche pour le développement (IRD) et le monde scientifique en général, notamment par la promotion des recherches menées par nos partenaires dans les pays du Sud.

*Jean-Paul Moatti, vous étiez, jusqu'à une date récente, président-directeur général de l'IRD. Mais d'abord, en tant que spécialiste en économie de la santé, vous connaissez bien les essais cliniques. Que pensez-vous des RCT à l'aune de votre propre expérience ?*

J'ai consacré la plupart de mes quarante années de carrière universitaire à collaborer étroitement avec des épidémiologistes et des biostatisticiens qui ont

longtemps considéré les évaluations randomisées comme un outil essentiel pour rationaliser la pratique médicale sur la base de preuves scientifiques, mais qui ont fini par exprimer certaines préoccupations sur la confiance excessive accordée aux protocoles randomisés.

On a longtemps pensé que les RCT constituaient la source idéale (le « *Gold Standard* », « l'étalon-or ») de données sur les effets des traitements médicaux. D'autres méthodes, comme les études de cohorte et les études cas-témoins, sont bien entendu utilisées lorsque la randomisation n'est pas possible pour des raisons éthiques ou pratiques, comme c'est souvent le cas des études sur les facteurs de risque environnementaux. Ces dernières années s'est néanmoins développé un intérêt croissant pour d'autres méthodes capables de fournir des preuves empiriques pour les stratégies les plus efficaces dans les domaines de la médecine et de la santé publique.

Des RCT bien conçues peuvent bien sûr prétendre à une forte validité interne dans la mesure où elles répartissent de manière équilibrée les facteurs connus et inconnus entre les groupes de contrôle et de traitement, limitant ainsi le risque d'introduire des facteurs de confusion dans l'identification d'un mécanisme causal. Pourtant, les experts en santé publique reconnaissent depuis longtemps les limites des RCT sur le plan de leur validité externe et de leur pertinence pour la prise de décision. En effet, il peut y avoir un certain nombre de raisons qui font qu'une RCT manque de validité externe, et la généralisation des résultats hors de la population étudiée peut s'avérer erronée. Les RCT ne portent généralement pas sur des périodes d'étude suffisamment longues ou sur des populations assez importantes pour pouvoir évaluer, dans la durée, les effets d'un traitement, comme dans le cas de l'impact des vaccins sur l'immunité à long terme d'une population, ou pour mettre en évidence des effets indésirables rares, mais graves du médicament, qui apparaissent seulement lors des phases de surveillance après la mise sur le marché au travers des dispositifs de pharmacovigilance. Les contraintes croissantes de coût et de temps pesant sur les RCT peuvent inciter à se fier à des marqueurs de substitution potentiellement mal corrélés au résultat clinique (mortalité ou morbidité grave) que l'on cherche en réalité à évaluer. Pour limiter la dimension de l'échantillon et garantir une puissance statistique suffisante, les RCT se concentrent souvent sur les groupes à haut risque, ce qui réduit fréquemment leur pertinence pour des populations cibles plus larges. En outre, il faut des années pour planifier, mettre en œuvre et analyser la plupart des RCT, ce qui limite leur capacité à suivre le rythme des innovations biomédicales et oblige à prendre des décisions concernant les nouveaux médicaments et les dispositifs médicaux alors que leur évaluation clinique est encore en cours. Les contraintes de temps associées aux RCT limitent également leur utilisation efficace pour les décisions de santé publique en cas de flambées épidémiques ou de crises sanitaires. En outre, des RCT différentes traitant d'un même sujet ont souvent produit des résultats contradictoires, notamment en ce qui concerne des questions clés sur l'efficacité des pratiques médicales. Cela a conduit au développement de méthodes dites de *méta-analyse*, qui combinent les résultats de différents essais afin de surmonter ces contradictions et de parvenir à des

conclusions fondées. Mais ces méthodes soulèvent elles-mêmes des problèmes statistiques complexes ainsi que d'identification des essais dont les données peuvent être agrégées sans créer une hétérogénéité excessive.

De manière générale, les experts en santé publique tendent désormais à considérer que les systèmes actuels de hiérarchisation des preuves empiriques mises en œuvre dans les mécanismes d'évaluation des pratiques médicales sont biaisés en faveur des RCT et qu'ils sont susceptibles d'écarter des données valides qui ne sont pas issues d'expérimentations randomisées.

Plus récemment, un nombre croissant de biostatisticiens ont reconnu que la randomisation ne constitue pas en soi une garantie absolue de validité interne. COOK (2018) recense par exemple 26 hypothèses susceptibles de biaiser les résultats des RCT malgré la randomisation, 22 d'entre elles étant liées à la validité interne : une différence préexistante à la sélection des groupes de traitement et de contrôle qui pourrait être confondue avec un effet de traitement ; la possibilité que l'attribution ait pu varier d'un groupe à un autre dans l'essai, rendant les résultats très sensibles au traitement des données manquantes (notamment liés aux abandons et perdus de vue en cours d'essai) ; un biais dans le choix du groupe de contrôle (par exemple lorsque l'innovation est comparée à un standard actuel de traitement qui n'est pas optimal) ; des changements de comportement induits par la participation à l'essai (par exemple, dans certains essais en double aveugle avec un groupe de contrôle prenant un placebo, les patients infectés par le VIH ont commencé à partager leurs pilules afin de garantir que tous les participants reçoivent « au moins » quelques médicaments efficaces), etc. La plupart de ces préoccupations tournent autour d'une question clé : si la randomisation, qu'elle soit volontaire ou involontaire, qu'elle intervienne *ex ante* ou *ex post* (dans le mode d'analyse de l'essai), ignore les informations préalables issues de la théorie et la prise en compte des covariables, alors elle est susceptible d'introduire des biais, et peut même s'avérer contraire à l'éthique, car elle expose inutilement les participants à un danger éventuel dans le cadre d'une expérience risquée.

Des modifications de la structure de base de l'évaluation randomisée ont été élaborées pour minimiser ce risque grâce à des mesures comme la stratification, l'allocation adaptative et la pré-randomisation, visant à prévenir la distorsion entre les groupes, qui peut venir, en dépit du tirage au sort, de facteurs pronostiques connus ou identifiés. Les modèles recourant à la méthode du cas unique (*Single Case Design* ou SCD) sont utilisés lorsqu'une variable d'intérêt dépendante peut être mesurée de manière répétée dans le temps entre deux points (au départ et pendant ou après l'intervention). Plutôt que de recourir à la randomisation d'un grand nombre de participants, les chercheurs utilisent un ordonnancement minutieux et planifié des conditions expérimentales afin d'améliorer la validité interne en excluant toute autre explication des effets du traitement. Toutes ces tentatives d'amélioration de la validité interne constituent *de facto* un aveu que la randomisation ne présente, en pratique, aucune supériorité statistique intrinsèque pour l'inférence causale.

*En mettant maintenant votre autre casquette, celle d'économétricien, vous connaissez sans doute bien les problématiques méthodologiques liées à l'inférence causale et les autres modèles de génération de groupes contrôle permettant d'identifier les facteurs causaux du phénomène d'intérêt.*

Avec ces évolutions et la tendance actuelle vers l'« expérimentation pragmatique », les experts en santé publique ont redécouvert un « vieil » argument économétrique bayésien remontant à FISHER (1926) et SAVAGE (1962), qui remet en question la croyance selon laquelle les effets de traitement moyens estimés à partir de RCT sont probablement plus proches de la vérité que ceux estimés par d'autres moyens. Le résultat visé par toute RCT réside dans l'écart de moyennes entre le groupe d'intervention et le groupe de contrôle, qui combine l'effet de traitement moyen estimé parmi les personnes traitées au terme d'erreur, qui lui-même traduit le déséquilibre généré aléatoirement parmi les effets nets issus d'autres causes. Les RCT offrent une base pour calculer l'ampleur de l'erreur, mais, comme indiqué précédemment, cela reste soumis à la condition qu'aucune corrélation avec des covariables ne soit intervenue avant ou après la randomisation. Que ce soit dans les RCT ou dans d'autres méthodes, la pertinence statistique sera menacée si la distribution des effets de traitement individuels dans la population étudiée est asymétrique. Dans sa synthèse des nombreuses publications économétriques sur le sujet, le prix Nobel d'économie Deaton soutient à juste titre que tout statut spécial accordé aux RCT est injustifié et conclut que « la méthode la plus à même de produire une bonne inférence causale dépend de ce que nous essayons de découvrir et de ce que nous savons déjà » (DEATON et CARTWRIGHT, 2018 : 2).

Les économètres connaissent d'autres méthodes conçues pour obtenir une inférence causale, comme l'appariement par scores de propension, les variables instrumentales, la modélisation économétrique et les réseaux bayésiens. Toutes les méthodes doivent bien entendu recourir à des groupes de contrôle pour garantir une comparaison appropriée avec le groupe d'intervention, mais le choix du modèle d'étude doit rester pragmatique et dépendre des problématiques en question.

Un autre prix Nobel d'économie, Heckman, met en lumière une autre source de scepticisme à l'égard des évaluations randomisées parmi les économètres, à savoir que les informations sur les effets de traitement moyens peuvent s'avérer peu utiles pour éclairer les politiques publiques, car elles ne prennent pas en compte les variations entre les bénéficiaires des interventions (HECKMAN et SMITH, 1995). Les impacts moyens peuvent être le facteur d'intérêt principal dans une évaluation comparant deux médicaments ou deux interventions très simples, mais lorsqu'il s'agit de politiques à composantes multiples, la possibilité de tirer des enseignements utiles d'une expérience spécifique nécessite plutôt d'identifier les raisons pour lesquelles certaines stratégies fonctionnent mieux que d'autres. Dans de tels cas, même une RCT réussie ne peut garantir que la relation causale établie se maintiendra dans d'autres contextes ou de manière générale. Il paraît clairement fallacieux de prétendre que la microfinance doit

être au cœur des efforts pour éliminer la pauvreté ou que les transferts monétaires conditionnels doivent être la priorité des politiques en matière de santé et d'éducation en s'appuyant sur un nombre limité d'évaluations randomisées réalisées dans ces domaines. Cela peut détourner l'attention des grandes politiques publiques destinées à réduire les inégalités ou à assurer la santé et l'éducation pour tous. Il est intéressant de constater que certains partisans des évaluations randomisées, tels que BANERJEE *et al.* (2015b : 52), arrivent à une conclusion similaire lorsqu'ils reconnaissent, par exemple, que « les adeptes du microcrédit ont peut-être [...] surestimé le potentiel des entreprises pour les pauvres, à la fois comme source de revenus et comme moyen d'autonomisation des femmes qui en sont les propriétaires ».

*Rémy Rioux, avez-vous observé des retombées opérationnelles des évaluations au niveau des projets que vous soutenez sur le terrain, et quelles évolutions avez-vous remarquées ou souhaiteriez-vous voir ?*

En général, nous constatons que la culture de l'évaluation favorise une culture de l'innovation. Les évaluations d'impact ont généré des retombées intéressantes dans la mesure où elles ont consolidé une culture de l'évaluation au sein de l'AFD et parmi nos partenaires des pays en développement.

En Mauritanie, par exemple, l'évaluation de l'impact d'un dispositif de protection sociale couvrant 40 % des femmes a montré que celui-ci augmentait significativement le recours aux soins et réduisait les inégalités, mais qu'il ne touchait pas les plus démunies et qu'il n'avait pas d'impact significatif sur la santé des mères et des enfants en raison de la dégradation de la qualité des soins dans les établissements (PHILIBERT *et al.*, 2017). En conséquence, nous avons complètement repensé la phase suivante du projet : action intégrale sur les différentes composantes de la qualité (ressources humaines, sang, médicaments et supervision) et opérationnalisation d'un mécanisme de gratuité pour les plus pauvres.

Les évaluations scientifiques d'impact ont également favorisé des innovations méthodologiques qui améliorent le suivi des projets. Par exemple, nous proposons désormais d'aider les responsables de projets et les partenaires à utiliser des données existantes dès l'instruction d'un projet pour mieux estimer les conditions de vie et l'accès aux services, analyser les dépenses des ménages, etc., ou encore à utiliser des images satellitaires pour suivre la productivité agricole, la déforestation, le développement urbain, etc. Le suivi continu des projets est également facilité par des outils numériques comme GeoPoppy, une application développée par l'Institut national de la recherche agronomique (Inra), que nous avons utilisée pour suivre l'agriculture en Côte d'Ivoire et que nous allons même faire évoluer en outil de renforcement des capacités au Bénin. Dans le même esprit, l'AFD collabore en Haïti et au Niger avec le Centre de recherches interdisciplinaires (CRI), fondé par François Taddei et Ariel Lindner, pour expérimenter et diffuser de nouvelles façons de mener des recherches et de mobiliser l'intelligence collective dans les domaines des sciences du vivant, de l'éducation et des technologies numériques.



Tous ces exemples montrent que, lorsque nous sommes à même de rendre compte des impacts de nos projets, nous apprenons à partager notre expérience, à mieux écouter nos bénéficiaires et, *in fine*, à innover pour eux et avec eux. En évaluant nos impacts, nous pouvons innover et montrer concrètement le retour sur investissement que génère la politique de développement, c'est-à-dire un investissement durable et solidaire.

*Jean-Paul Moatti, vous êtes aussi engagé dans la contribution aux politiques publiques. Outre votre parcours académique, vous avez été membre du groupe d'experts des Nations unies chargé du rapport mondial sur le développement durable (Global Sustainable Development Report – GSDR). À ce titre, vous avez participé à la rédaction du premier rapport quadriennal d'évaluation des ODD, qui ont été adoptés par tous les États membres des Nations unies en septembre 2015 et qui fixent l'Agenda international 2030 pour le développement. Nous assistons dans ce cadre à l'émergence du concept de « science de la durabilité ». Soutenez-vous ce concept et comment les expérimentations randomisées peuvent-elles, ou non, contribuer à une recherche efficace pour le développement durable ?*

Même si les 17 ODD restent le fruit de nombreux compromis entre les gouvernements et entre des intérêts contradictoires, ce programme de transformation ambitieux a très largement bénéficié de l'émergence de ce que l'on appelle aujourd'hui la « science de la durabilité ». L'Académie nationale des sciences des États-Unis, qui a commencé à promouvoir la science de la durabilité en 2000, la définit comme un « domaine de recherche émergent qui traite des interactions entre les systèmes naturels et sociaux, et de la manière dont ces interactions affectent le défi de la durabilité : satisfaire les besoins des générations actuelles et futures tout en réduisant sensiblement la pauvreté et en préservant les écosystèmes qui entretiennent la vie sur la planète » (KATES, 2011). Parce qu'elle est axée sur les problèmes, cette nouvelle approche scientifique est par nature interdisciplinaire et se prête à la coconstruction de programmes de recherche avec les communautés concernées. Elle s'attache à identifier les chaînes de causalité complexes qui génèrent les principales problématiques environnementales et sociales menaçant l'avenir de la Terre, et à proposer des solutions pour réduire le risque d'incohérence dans la mise en œuvre des ODD tout en maximisant les synergies positives qui existent entre eux : comment augmenter la productivité agricole pour garantir la sécurité alimentaire d'une population mondiale, censée augmenter de deux milliards de personnes d'ici 2050, tout en réduisant les intrants chimiques afin de limiter l'impact environnemental et le gaspillage des ressources ? Comment promouvoir une croissance durable pour éliminer l'extrême pauvreté sans accroître les inégalités à l'intérieur des pays ? Les enjeux sont multiples.

L'économie du développement devrait jouer un rôle de premier plan dans cette science de la durabilité, car elle mobilise des compétences et des connaissances essentielles pour la transposition en politiques publiques efficaces de faits et de preuves empiriques, issues d'une gamme étendue de disciplines allant des

sciences naturelles aux sciences sociales, afin de les adapter à des contextes sociaux, environnementaux et économiques hétérogènes. Les RCT ne sont bien souvent pas adaptées à ce champ interdisciplinaire essentiel, car l'extrapolation et la généralisation de leurs résultats nécessitent un ensemble d'informations complémentaires qui doivent provenir d'autres sources. Surestimer leur rôle, en ignorant les limites de l'approche randomisée et en exagérant leurs mérites auprès des décideurs, pourrait compromettre la contribution de la science économique à la transformation des modèles de développement actuels vers la durabilité. Cependant, ce serait également une erreur de sous-estimer le fait que les études randomisées peuvent s'avérer extrêmement utiles, le cas échéant, pour déterminer quelles sont les meilleures pratiques, parmi différentes modalités d'intervention, pour atteindre les ODD et pour produire des arguments puissants en faveur de politiques de changement social et de transition écologique basées sur des preuves empiriques.

*Une dernière question pour tous les deux : comment des institutions comme l'AFD et l'IRD peuvent-elles coordonner leurs efforts pour que les pays du Sud bénéficient d'actions de recherche pertinentes qui soient utiles et utilisées pour l'élaboration et la mise en œuvre des politiques ?*

*Rémy Rioux* : Nous pensons que tout le monde gagne lorsque la recherche est menée en partenariat avec les décideurs et la société civile. Ces partenariats doivent préserver l'indépendance et la rigueur de la recherche et favoriser une fertilisation mutuelle afin que la production intellectuelle soit pertinente, appropriée, et qu'elle contribue au progrès de nos sociétés. Cette implication doit se manifester à toutes les étapes du cycle de la production scientifique, depuis la conception et le cadrage de la recherche, à sa mise en œuvre et à la formalisation des connaissances et leur diffusion.

Le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) est un bon exemple de cette dynamique. Il est essentiel de combiner l'action contre les changements climatiques à des politiques visant à réduire les inégalités et à renforcer le lien social dans nos sociétés pour garantir que la transition environnementale, si vitale aujourd'hui, soit également socialement durable. Cette recherche dont nous avons besoin devrait être implantée dans le Sud, avec le soutien et l'accompagnement de centres d'excellence du Nord lorsque c'est pertinent. « Le Sud inspirera le Nord », pour citer le président d'Unicef France, Jean-Marie Dru (communication personnelle).

À cet égard, l'IRD est un partenaire précieux, car il est reconnu pour son excellence scientifique et ses liens étroits avec les équipes universitaires des pays du Sud. Sa mission est en outre entièrement dédiée aux pays en développement, et tous ses travaux sont menés en partenariat et pour renforcer les capacités de recherche dans le Sud. Nos deux institutions sont engagées dans une démarche conjointe depuis 2012 en matière de recherche sur le développement durable. Notre collaboration s'appuie sur l'approche de la science de la durabilité, promouvant l'interdisciplinarité et le rapprochement entre connaissances scientifiques et les savoirs des autres acteurs du développement.

Dans la même veine, l'International Development Finance Club (IDFC), que je préside depuis 2017, œuvre pour que les pays du Sud bénéficient de travaux de recherche pour élaborer et mettre en œuvre des politiques publiques. Le Club a, de fait, présenté récemment un rapport novateur, rédigé par les groupes de réflexion indépendants CPI (Climate Policy Initiative) et I4CE (Institute for Climate Economics), qui fournit un cadre solide, utilisable par les 26 banques de développement nationales et régionales membres de l'IDFC – dont beaucoup sont basées dans les pays du Sud –, et par la communauté financière dans son ensemble, afin d'aligner les stratégies de toutes les institutions financières avec les objectifs de l'accord de Paris aux niveaux national, stratégique et opérationnel. L'évaluation des projets fait également partie des sujets discutés au sein du Club afin d'identifier les questions soulevées par l'évaluation des actions pour le climat par les organisations internationales et les bailleurs de fonds, et notamment l'examen des défis méthodologiques liés à la mesure des impacts des programmes de développement concernant le changement climatique. Dans les mois à venir, le lancement de la plateforme électronique de l'IDFC permettra au Club de favoriser les échanges entre les experts et de promouvoir le partage des connaissances et des bonnes pratiques.

*Jean-Paul Moatti* : L'IRD a derrière lui une longue histoire, puisqu'il a fêté ses 75 ans en 2019. Il travaille dans plus de 50 pays en développement. L'AFD bénéficie de l'expertise des chercheurs de l'IRD depuis de nombreuses années. Paradoxalement toutefois, bien que la France reste le seul pays avancé à disposer, avec l'IRD, d'un organisme public de recherche interdisciplinaire, dont la mission unique est la coopération scientifique équitable avec les collègues, chercheurs et universitaires, des pays en développement, nous avons travaillé moins fréquemment et moins systématiquement avec l'AFD que d'autres agences et banques de développement – comme l'USAID et le DFID britannique – ne l'ont fait avec leurs chercheurs nationaux. L'une des raisons est que, jusqu'à une date récente, le financement de la recherche ne faisait pas partie de la mission de l'AFD, les contributions scientifiques devant donc emprunter des canaux d'expertise contractuelle qui n'étaient pas toujours bien adaptés aux projets de recherche. Mais la situation évolue rapidement, alors que les ODD et l'action contre le changement climatique deviennent des axes de travail communs à l'AFD et à l'IRD, et que le rôle majeur de la diplomatie scientifique dans le développement durable est de plus en plus reconnu (voir le GIEC pour le climat et l'Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [IPBES] pour la biodiversité et, bien sûr, le GSDR).

Premièrement, et avant tout, l'AFD soutient maintenant des programmes de renforcement des capacités universitaires et de recherche dans les pays en développement, comme on peut le voir en Côte d'Ivoire, où l'AFD intervient, conjointement avec le programme français de désendettement et dans sa collaboration avec la Banque mondiale, pour soutenir les Centres d'excellence africains (African Center of Excellence – ACE) dans les domaines de la recherche, notamment en Afrique francophone. De nombreux partenaires africains du projet ACE de l'IRD reçoivent aujourd'hui l'aide de l'AFD. Deuxièmement, le nouvel

accord global signé début 2019 entre l'AFD et l'IRD vise à associer plus étroitement, dans les pays, les initiatives décentralisées menées par l'AFD avec des chercheurs aux grands programmes codéveloppés par les deux organisations. Ces programmes incluront des projets d'évaluation scientifique et de développement financés par l'AFD, qui pourraient s'avérer riches d'enseignements et permettre un certain niveau de généralisation à l'échelle internationale. L'élaboration d'un protocole d'évaluation expérimental ou quasi expérimental *ex ante*, randomisé ou non, et la mise en place simultanée d'un projet de développement durable augmentent les chances de réussite, tant pour l'évaluation scientifique que pour le projet lui-même. Une collaboration plus étroite avec l'AFD est un excellent moyen de promouvoir les changements requis par la science de la durabilité en matière de pratiques de recherche, ceci incluant la coconstruction de programmes de recherche avec les communautés et les groupes de population vulnérables qui sont directement concernés.

# Expérimentations aléatoires dans le champ du développement

Une perspective critique

Éditeurs scientifiques

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

**IRD Éditions**

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT

Collection Synthèses

Marseille, 2022

Ce texte a fait l'objet d'une première édition en anglais :

BÉDÉCARRATS F., GUÉRIN I., ROUBAUD F. (eds), 2020, Randomized Control Trials in the Field of Development : a Critical Perspective, Oxford, Oxford University Press, 448 p.

<https://global.oup.com/academic/product/randomized-control-trials-in-the-field-of-development-9780198865360>

Photo 1<sup>re</sup> et 4<sup>e</sup> de couverture

© Ermell – Paul Klee, *Klippen am Meer*, 1931, Musée Lenbachhaus, Munich.

[https://upload.wikimedia.org/wikipedia/commons/6/69/Klee\\_Klippen\\_am\\_Meer\\_1290074.jpg](https://upload.wikimedia.org/wikipedia/commons/6/69/Klee_Klippen_am_Meer_1290074.jpg)

### **Coordination éditoriale**

IRD/Romain Costa

### **Préparation éditoriale**

Marie-Laure Portal (11)

### **Mise en page**

Desk (53)

### **Maquette de couverture**

IRD/Michelle Saint-Léger

### **Maquette intérieure**

IRD/Pierre Lopez



© IRD, 2022

Certains droits réservés. Il s'agit d'une publication en open access, disponible en ligne et distribuée sous les termes de l'attribution Creative commons, non commerciale, non modifiable 4.0 licence internationale (CC BY-NC-ND 4.0), dont une copie est disponible à cette adresse :

<http://creativecommons.org/licences/by-nc-nd/4.0/>

ISBN papier : 978-2-7099-2947-9

ISSN : 2431-7128

ISBN PDF : 978-2-7099-2948-6

ISBN epub : 978-2-7099-2949-3