

Retranscription de l'entretien avec Florent Bédécarrats

A peu près tous les acteurs semblent être d'accord pour dire qu'une complémentarité entre les différentes méthodes (quantitatives, qualitatives, opérationnelles...) est nécessaire. Même lorsque nous nous sommes rendus au J-PAL, il nous a été expliqué que les chercheurs du J-PAL essaient de mettre en place une telle complémentarité en introduisant d'autres méthodes plus qualitatives sur le terrain pour mieux comprendre les effets causals et ne pas simplement avoir une « boîte noire ». Selon vous, à quelle hauteur situe-t-on cette complémentarité et comment fait-on pour la mettre en place en pratique au niveau de l'expérience ?

Parmi les tenants des approches quantitatives il existe une ambiguïté sur ce qu'on entend par « qualitatif ». Pour des quantitativistes, le qualitatif se résume, le plus souvent, à une variable de « perception » dans le questionnaire. Certains sont friands d'approches psychologisantes : faire des jeux, poser aux gens des questions du type « est-ce que vous préféreriez que je vous donne 10 francs tous les jours ou 100 francs dans 8 jours ? » etc... Ce sont parfois des modalités un peu schématiques ou, en tout cas, avec une approche très codifiée du qualitatif qui est souvent perçue par les qualitativistes comme incomplète et partielle par rapport à ce qu'est, selon eux, du « vrai » qualitatif. Il y a donc un malentendu sur ce qu'est le qualitatif.

Pour d'autres quantitativistes, le qualitatif va être une manière d'aller un peu plus loin et de dire : « on va faire en sorte de faire quelques observations, entretiens, focus groupe, avant de finaliser notre questionnaire pour éventuellement ajuster les hypothèses, le dispositif, vérifier que le contrefactuel est vraiment valide et puis s'assurer d'avoir les bonnes variables ». Cela peut donc être fait éventuellement en amont, mais également en aval, au moment d'interpréter les résultats : aller refaire un petit tour de qualitatif pour les contextualiser, les réinterpréter etc. C'est un peu une approche médiane.

Il y a une dernière approche, qui est la plus qualitative : les tenants du qualitatif expliquent qu'il faut arrêter de considérer cette approche un subsidiaire au quantitatif. Celle-ci devient alors un mode d'administration de la preuve à part entière. C'est pourquoi les qualitativistes argumentent en disant : « vous, les quantitativistes, ne gardez comme critère de rigueur que la taille de votre échantillon, son mode de constitution aléatoire et éventuellement le mode de constitution du contrefactuel de manière aléatoire. Mais il y a d'autres critères

de rigueur et notamment des critères de triangulation de l'information, d'analyse et objectivation des représentations, et enfin de distanciation par rapport aux représentations y compris les miennes en tant que chercheur, mais aussi tout ce qui est la pratique du terrain ». Jean-Pierre Olivier de Sardan a écrit un bouquin sur la rigueur dans le qualitatif, que je vous invite à lire, qui fait la synthèse, en tout cas du point de vue de la socio-anthropologie du développement, sur ce qu'est la rigueur.

C'est paradoxal, car c'est un peu un poncif de dire qu'il faut articuler quantitatif et qualitatif : tout le monde est d'accord pour le souhaiter et le prévoir, mais finalement, cela ne se fait jamais.

Pourquoi, s'ils sont tous d'accords, nous dit-on qu'il y a une controverse là-dessus ? Où chacun veut-il placer le curseur et qu'est-ce qui pose problème mise à part la définition ?

C'est en fait une question de pouvoir et d'argent. En effet, comme on parle de projets de recherche, il y a des enjeux de rétribution pour les chercheurs. Dans le domaine de la recherche, ces rétributions ne sont pas purement financières mais aussi en termes de prestige. Par exemple la question de savoir quelle personne signe en première les publications, ou dans quelle revue est-ce que l'article sera publié. Or les qualitatifs ne publient pas dans les mêmes revues que les quantitativistes. Il y a donc beaucoup d'enjeux de valorisation de la recherche. Ceux-ci ont des conséquences tout à fait tangibles pour les chercheurs, par exemple : « est-ce qu'on va me payer un post-doc ? », « est-ce que je vais pouvoir m'en servir pour financer un projet de recherche ? » etc...

Les postures épistémologiques et les démarches fondamentales du qualitatif et du quantitatif sont très différentes : hypothético-déductives pour le quantitatif ou alors plutôt inductive ou abductive du côté du qualitatif. Il y a une manière de représenter la vérité et de démontrer qui n'est vraiment pas la même dans un camp et dans l'autre.

Un des problèmes réside dans le fait qu'il y a des enjeux factuels quand on a un budget pour une évaluation : par exemple les questions de savoir « qui est financé ? », combien de temps prend l'expérience... Généralement la partie quantitative est la plus coûteuse. En tout cas c'est celle qui concentre le plus d'argent et qui pose le plus de contraintes opérationnelles dans les enquêtes. C'est également celle qui va le plus souvent être le critère de qualité déterminant pour pouvoir publier dans une revue A+ en économie. Elle prend donc le pas sur les autres. Par ailleurs, les questions se posent de savoir qui aura, contractuellement, le rôle de pilote dans l'évaluation, qui sera la

personne maître d'œuvre, qui décidera éventuellement de déléguer, de sous-traiter. En général, c'est souvent le quantitativiste qui a le rôle de pilote.

Vous parlez de visions différentes de la vérité et vous expliquez dans votre article qu'une partie du succès vient notamment de l'apparente scientificité que présente l'écriture mathématique du quantitatif par rapport au qualitatif. Par ailleurs, la « vérité » telle quelle est présentée par les RCTs est souvent décrite comme plus accessible. Cela participe-t-il à donner le pas aux RCTs ?

Oui, il y a vraiment un argument de simplicité qui est utilisé, mais c'est seulement un argument. Je vous invite à regarder les évaluations randomisées que l'on a financées à l'AFD : l'une au Maroc et l'autre au Cambodge. Celle du Maroc avec Esther Duflo, Bruno Crépon, Florencia Devoto et William Pariente. Et celle du Cambodge avec David Levine et Rachel Polimeni, de l'université de Berkley, appartenant à un laboratoire qui s'appelle CEGA.

Il y a des « *cautionary tale* » - anecdotes ou formules marquantes - dans le domaine de l'évaluation de l'impact. L'argument de la simplicité est très utilisé mais quand on regarde comment ça se passe sur le terrain, par exemple au niveau des deux évaluations que l'on a effectuées, les résultats sont tellement compliqués à analyser, on a des couches et des couches d'économétrie par-dessus et finalement l'argument de la simplicité est beaucoup moins visible. On est sûr des interprétations au contraire très complexes.

Ce n'est pas uniquement une question d'avoir l'air plus accessible. Il y a une valeur sociale du chiffre et de la démonstration statistique, au sens d'administration de la preuve, qui est plus forte pour les méthodes quantitatives. Dans le monde du développement il y a actuellement un très fort courant en faveur des « *evidence-based policies* ». L'idée est de faire de la preuve statistique l'alpha à l'oméga de la décision politique. C'est reconnu, valorisé, revendiqué par les décideurs.

Selon moi, il ne faut pas seulement voir ce débat en terme d'opposition entre quantitatif et qualitatif : il y a un troisième pôle qui est celui des opérationnels – les personnes qui travaillent sur le terrain, comme le personnel des ONG par exemple. Je ne sais pas si vous avez lu le lancement de cette grande controverse initiée par le rapport « *When will we ever learn ?* » du CGDEV en 2006, auquel ont pris part tout ce que l'économie quantitative compte de plus éminent. Ils se sont mis d'accord et ont co-signé un rapport qui explique ce que beaucoup de *randomistas* répètent : le fait qu'on gaspille de l'argent depuis des dizaines d'années sans savoir ce qui marche et ce qui ne marche

pas. Duflo fait le parallèle avec l'évolution depuis le Moyen-âge où on traitait les gens avec des sangsues, jusqu'à l'âge de raison aujourd'hui: l'idée est que grâce aux nouvelles méthodes, on va pouvoir sortir le monde de l'ignorance.

L'un des problèmes c'est qu'il y a quand même une négation des savoirs et de la réflexivité des opérationnels. Celle-ci s'appuyait beaucoup au départ sur un champ de vision assez restreint où la critique s'adressait essentiellement aux théoriciens de l'économie, par rapport aux grands paradigmes : ajustements structurels versus investissements. Mais il y a aujourd'hui une vraie négation des savoirs pratiques et de l'expérience pratique alors que les méthodes quantitatives concernent justement des savoirs pratiques (par l'analyse de micro-dispositifs comme : « comment est-ce qu'on rémunère des professeurs ? », « comment on les forme ? », « comment est-ce qu'on construit leur travail ? », « quel type de garantie pour du crédit ? »). On est vraiment sur des micro-dispositifs donc il y a une impression forte de négation de ce que font des gens qui sont des ingénieurs : ces gens-là ont une vraie culture dans leur domaine, et ont l'impression d'accumuler du savoir et de l'expertise technique depuis des décennies, et on vient leur dire que « ça, ce n'est pas de la science ».

Par rapport à l'idée de négation, on parle beaucoup aussi de la négation de tous les écrits dans les domaines de la sociologie ou de l'anthropologie par exemple. Est-ce que vous partagez l'avis de personnes comme M. Nubukpo qui explique dans son article d'Alternatives Economiques que les *randomistas* font fi de toute la littérature sur les sujets qu'ils traitent pour apporter des découvertes qui n'en sont finalement pas vraiment ? De même, vous disiez qu'il est intéressant de faire du qualitatif en amont pour savoir dans quel cadre on va travailler. Voyez-vous la volonté des *randomistas* de travailler sans cadre, de façon a-théorique et très empirique, comme un frein pour la complémentarité avec le qualitatif.

Complètement. Je ne sais pas si vous l'avez perçu, mais le rôle institutionnel des acteurs est important pour comprendre leur position dans la controverse. Par exemple, j'avais personnellement une liberté de parole très libre et une posture très critique de la randomisation il y a quelques mois et je l'ai beaucoup moins maintenant que je travaille à l'AFD car ma fonction implique de travailler avec l'ensemble des acteurs du champ et de chercher à construire de ponts. De ce point de vue-là, les propos de Kako Nubukpo peuvent être perçus comme un peu violents ! Cependant, sur le constat sur le fond, il a raison sur l'idée que ce courant tend à « faire du passé table rase » : d'une part du point de vue de la théorie sociologique, anthropologique voir économique plus hétérodoxe qui n'existe pas dans le paysage et n'est pas citée, mais

également du point de vue des apprentissages opérationnels. J'ai écrit un article sur les évaluations d'impact dans le domaine de la micro-finance. On a véritablement l'impression qu'on réinvente l'eau chaude dans le sens où on voit depuis 2006-2007, la grosse vague arrivant surtout en 2010-2011, des évaluations d'impacts qui présentent comme des découvertes inédites des choses qui ont déjà été étudiées, re-étudiées et re-re-étudiées il y a quinze ou vingt ans dans le domaine de la microfinance. Un bon exemple concerne le type de garantie avec lequel on propose un prêt et l'implication que cela a sur le remboursement ou le profil des gens qui y ont accès : il n'y a pas une, deux ou trois mais des centaines de références de travaux qui n'étaient certes pas expérimentaux mais empiriques. En Bolivie, par exemple, des travaux ont été menés sur quasiment dix ans sur de nombreuses d'institutions, des milliers d'emprunts ont été analysés... On a l'impression qu'on découvre aujourd'hui la garantie solidaire, qui permet de résoudre l'asymétrie de l'information et permet à des gens qui n'avaient pas accès au crédit, d'y avoir accès. Il y a un peu une posture autoréférentielle ou a-référentielle qui ne facilite pas le dialogue avec d'autres démarches ou d'autres théories.

Sur l'aspect autoréférentiel justement, vous expliquez que tous les comités de lecture des plus grandes revues sont « trustées » par des partisans du quantitatif et des RCTs. Est-ce qu'il n'y a pas un cercle vicieux qui se crée puisqu'il y a eu un effet de mode qui a fait que les RCTs étaient en vogue donc on a mis les bons chercheurs qui faisaient des RCTs aux comités de lecture et le mécanisme s'est autoalimenté ?

Oui en effet, bien sûr. Un point intéressant selon moi, réside dans le fait que les RCTs sont arrivés avec la volonté d'arrêter ce type de cycle. La dynamique que vous décrivez-là favorise les grands mouvements de balancier : on pousse une logique, un paradigme à son extrême jusqu'à ce qu'il explose et qu'on reparte dans une direction diamétralement opposée. Les RCTs s'étaient affichés comme une manière de dépasser ça, de dire « nous on arrête l'idéologie, on va enfin être empiriques, objectifs, pour dépasser un débat stérile ». En fin de compte, on assiste actuellement au même phénomène de création d'un système autoréférentiel qui a tendance à s'autoalimenter jusqu'à se caricaturer à l'extrême et se retrouver face à ses contradictions. C'est difficile à dire parce que cela me concernait également avant d'avoir maintenant une position qui doit rester neutre et au-dessus de ces débats-là. C'est un peu un jeu de rôle, c'est comme un ping-pong : les autres finissent par être très agressifs aussi, à attaquer et à caricaturer. C'est quelque chose de systémique qui fait que les gens se polarisent. Il y a un phénomène véritablement de polarisation à la fois méthodologique, épistémologique mais aussi de chapelles. C'est d'ailleurs

parfois un peu épuisant. Je le présente comme ça mais, bien sûr, on essaie de dépasser cette logique. Il faut voir qu'il y a des médiateurs dans chacun des camps. C'est là où l'approche d'analyse des controverses est vraiment intéressante : on a toujours un peu l'impression d'être aux prises avec un débat unique mais on se rend compte que ce sont des schémas récurrents. Ça dure depuis les querelles théologiques, ce sont toujours les mêmes dynamiques de polarisation.

Par rapport à ces modes, il y a de gros enjeux sur la simplification du discours, sur la création de comptes moraux, de « *cautionary tales* ». Ces anecdotes ou formules qui ressortent partout afin d'expliquer les choses simplement en deux ou trois termes favorisent aussi la polarisation. Lorsque vous êtes spécialiste de la question vous comprenez que c'est plus complexe. Cependant, dès lors qu'on vous fait intervenir en tant que spécialiste face à des gens qui ne le sont pas, vous reprenez les arguments schématiques pour marquer les esprits et faire passer votre point de vue. Cette démarche de simplification est inhérente à l'augmentation de visibilité de la controverse, bien qu'on ne parle pas non plus ici de l'espace public. Elle a tendance à favoriser le côté caricatural.

Une autre chose est très importante pour moi, c'est la personnification.

On observe que l'aspect médiatique est très important et qu'une des forces des RCTs réside dans la vulgarisation.

Oui, et c'est aussi ce genre de choses qui crispent un peu les opposants, par exemple des titres d'articles comme : « Des neurones contre la misère ».

La personnification a joué un rôle important et est utilisée comme un argument de vente. C'est pourtant un énorme risque de réputation : un secteur entier, un courant de recherche extrêmement composite, met sur un piédestal une personnalité, et lie ainsi son destin à la crédibilité de cette personne-là. Je pense que Mme Duflo est quelqu'un de très compétent mais regardez le nombre d'évaluations qu'elle suit ! C'est la même chose pour Dean Karlan. C'est vertigineux. Quand on connaît l'ampleur des activités que ça requiert ! Personnellement, je m'interroge un peu sur l'emballage de la machine et sur les difficultés que ça va inévitablement susciter. Par ailleurs, concernant le côté médiatique, les médias sont extrêmement versatiles, ils adorent avoir une personnalité à mettre en exergue et ils adorent la vouer aux gémonies un peu plus tard...

On observe un peu un nouveau « scientific business model » avec un laboratoire comme la J-PAL : ce modèle est-il amené à se développer dans le monde recherche ?

Ce n'est pas vraiment nouveau, ça a toujours été un peu le cas. Les prix Nobel sont un peu dans cette logique, et les universités américaines aussi avec par exemple les petites vidéos en ligne du MIT, de Stanford. Il y a une véritable mise en scène de la recherche. Les différents business model viennent aussi de la diversité des modèles culturels : la manière de construire les icônes n'est pas la même selon les bases culturelles. En France il y a une notion de noblesse d'une recherche un peu éthérée. Les RCTs sont un peu à cheval entre les deux côtés de l'atlantique et jouent là-dessus pour construire une iconographie. Ce modèle n'est donc, pour moi, pas propre aux RCTs ni à la science : il participe d'une dynamique institutionnelle et culturelle occidentale du 20^{ème} siècle.

Pouvez-vous nous parler du rôle de l'AFD dans la controverse ?

Les phases ont été les suivantes :

L'AFD commence à s'intéresser aux évaluations d'impact au début des années 2000, en même temps que cela reste en général en marge (microfinance à Madagascar, culture en Guinée...). En 2006, un changement se produit sous la direction de Jean Michel Severino, anciennement à la banque mondiale. C'est quelqu'un qui a réellement positionné l'AFD, qui en a fait ce qu'elle est aujourd'hui et qui sur le plan de la recherche a mis l'accent sur la production de connaissance pour faire un peu le parallèle avec ce que faisait la banque mondiale à cette époque : « Knowledge Bank ». L'idée est ne pas seulement participer au financement mais également aux idées. L'évaluation est logée au département de la recherche avec une vocation première de production de connaissance.

Ce changement de direction coïncide avec le moment où l'on veut promouvoir les évaluations d'impact de manière plus systématique et plus organisée. Il faut que vous lisiez le rapport « *When will we ever learn* », qui prône une évaluation rigoureuse des programmes de développement au moyen de RCT. On se rend compte qu'il y a un peu une généalogie : l'année précédente, Angus Deaton et un groupe d'évaluateurs notamment des gens qui appartiendront au J-Pal par la suite, ont fait un rapport extrêmement sévère sur l'activité de recherche de la Banque Mondiale. Ce rapport d'évaluation par le *Groupe d'évaluation indépendante (IEG)* de la Banque Mondiale sauf erreur a remis en question des approches jugées trop guidées par de l'idéologie économique, et trop peu rigoureuses.

Dans ce contexte-là, l'AFD qui veut avoir une division d'évaluation au sein de la recherche décide de faire de l'évaluation d'impact. A cette époque se créent plusieurs fonds, dans une sorte de mouvement international, pour financer des évaluations d'impact. Plutôt que d'abonder l'un de ces fonds, l'AFD souhaite

pouvoir juger sur pièces et entreprend de commanditer deux évaluations d'impact en direct en les suivant de très près pour se faire son propre avis sur cette nouvelle méthode. L'idée est de pouvoir ensuite financer ces méthodes innovantes, et éventuellement aussi pour favoriser l'apprentissage en interne.

La volonté était donc avant tout de travailler avec des laboratoires prestigieux qui comptaient dans le domaine de la randomisation. La question de savoir à quel programme en particulier on allait choisir pour appliquer ces méthodes est venue dans un deuxième temps. On a coutume de dire dans l'évaluation d'impact qu'il faut choisir la méthode en fonction de ce que l'on veut évaluer. Là, l'inverse s'est produit. Ces deux évaluations se sont révélées être des processus complexes et très mitigés. L'étude au Maroc s'est déroulée entre 2006 et 2009-2010 car elle a rencontré des difficultés, et celle au Cambodge entre 2009-2012. Pour plusieurs raisons, ces évaluations n'ont pas si bien marché que ça. Un bilan assez critique en a été tiré et a été largement communiqué par l'AFD. L'agence a donc pris part non seulement à la promotion des RCTs avec la chaire au collège de France et la mise en place de deux évaluations mais a aussi fait part de ses critiques par la suite. Le bilan, mitigé, explique que certains aspects de la méthode sont intéressants mais que cette dernière s'applique à un champ trop restreint des projets, qu'elle permet de montrer des choses assez basiques et simples et également qu'elle ne permet pas d'aller très loin dans le temps et dans la chaîne des résultats. Par ailleurs, l'utilité de ces résultats, c'est-à-dire ce qu'on peut en tirer pour guider les politiques, n'est pas évidente : c'est le débat sur la validité externe des évaluations, à savoir la représentativité des résultats de l'expérience.

Par la suite, l'AFD a eu la volonté de continuer l'évaluation avec des méthodes un peu plus mixtes. L'avènement de ces choses-là s'est déroulé en 2012 avec une conférence EUDN à Paris dans laquelle l'AFD a présenté les résultats. Par la suite, il y a eu une nouvelle direction, moins favorable à la recherche et moins favorable à l'évaluation. A titre d'exemple, un gros projet d'évaluation d'impact sur l'eau n'a pas été accepté en République Démocratique du Congo. Ce refus peut vraiment être relié à l'expérience précédente mitigée de l'AFD. Le directeur a rétorqué que les évaluations d'impact étaient trop chères et trop complexes. S'en est suivi un creux entre 2011 et 2013-2014 sur les évaluations d'impact. On est en train de repartir aujourd'hui parce qu'il y a une volonté plus franche de la nouvelle direction de faire de l'évaluation avec des approches plus mixtes, bien qu'on fasse encore un peu de randomisation.

Travaillez-vous avec le J-PAL pour la mise en place de ces évaluations avec des méthodes mixtes ?

On est encore en discussion avec eux. A terme on aimerait participer à des approches mixtes avec eux. On est en train de revoir notre manière de faire. En effet, jusqu'à présent on a eu une approche un peu opportuniste : les chercheurs ou les chefs de projets nous expliquent : « j'ai un super projet, il faut l'évaluer ». L'évaluation ne s'inscrit donc pas toujours dans une réflexion plus globale sur les questions cruciales pour définir la politique de l'AFD dans les 5 ans à venir. On aimerait désormais intervenir plus en amont dans le processus de susciter des évaluations sur les thèmes qui nous paraissent vraiment stratégiques.

Certains fonds multilatéraux ont maintenant une approche plus ouverte sur cet aspect mais les vraies questions sont « qui finance ? » et « qui gouverne ? » l'évaluation.

L'une des grandes questions est « quels sont les enjeux ? » et « qu'est ce qui existe ? ». Tout le monde n'a pas la même définition de ce qui existe : est-ce qu'il faut que ce soit randomisé pour que ce soit considéré comme existant ou est-ce que d'autres modes de production de la connaissance sont valides ? A-t-on forcément besoin de refaire des évaluations très chères sur des thèmes pour lesquels on a déjà des réponses par d'autres méthodes ?

Pour vous, s'il doit y avoir une médiation, qui doit la porter ? Vous, en tant que financeur à l'AFD par exemple, ou directement les laboratoires eux-mêmes.

Pour l'instant ce qui est le plus faisable c'est qu'une telle médiation se fasse au niveau des bailleurs de fonds ou des instances de coordination entre bailleurs parce que le financement reste quand même le nerf de la guerre ! Ce sont soit des bailleurs qui financent directement (l'AFD, le DIFID) soit, pour des montants plus importants, les multilatéraux comme le 3IE. Ces derniers sont des gens de plus en plus ouverts : ils font des papiers très centrés sur les méthodes mixtes. Pour moi il existe une limite liée au fait qu'on reste quand même très focalisés sur l'évaluation d'impact. Le financement est encore très contrôlé par un petit nombre de bailleurs qui sont les plus gros contributeurs : Gates, le DIFID... L'approche est très axée sur le quantitatif et le qualitatif est toujours vu comme une manière d'apporter un petit plus. En même temps, il faut reconnaître que c'est tellement compliqué déjà de monter ces expérimentations-là... Elles peuvent paraître simples quand on les présente sur des petits *papers* mais ce sont des usines à gaz ! Il y a tellement d'acteurs qui interviennent. Du coup, une fois que la machine est lancée, l'enquête qualitative est très compliquée à mettre en place. On peut comprendre que pour des principes de réalité on ait envie d'avancer. Pour moi, des instances de financement de la recherche comme l'ANR en France devraient avoir un rôle pour promouvoir l'innovation

méthodologique et la pluridisciplinarité, mais elles ne s'intéressent pas trop à l'évaluation. Une des différences est qu'alors que dans les pays anglo-saxons, l'évaluation est une discipline à part entière, ce n'est pas le cas en France. Il y existe peut-être un ou deux masters, mais ils ne font pas légion et ils ont parfois du mal à exister et à être reconnus.

Les enjeux sont donc principalement le financement, la formation, et les publications dans les revues. Ce sont les publications et les comités de lecture qui créent la légitimité dans le milieu scientifique.

Comités qui sont déjà trustés par les randomistas ?

C'est difficile à dire car ils sont anonymes. ! Pour vraiment l'objectiver, il faudrait être un peu plus quantitatif et ça pourrait être intéressant de voir comment ça a circulé au sein de ces comités et dans quelle mesure ils regroupent majoritairement ou non des partisans des méthodes quantitatives.

On parle beaucoup des essais randomisés comme étant potentiellement destructeurs en occupant toute la place mais également du fait qu'ils encourageraient les évaluations en général. Qu'en est-il selon vous ?

C'est difficile à estimer, mais pour moi les effets d'éviction sur les ressources priment. Les enquêtes menées de manière ad hoc, comme les RCTS, impliquent qu'une partie des budgets ne vont pas chez les INS (Institut national de la statistique). Or ces enquêtes ne servent que pour leur propre finalité. En privilégiant d'autres méthodes, on pourrait probablement avoir une statistique publique qui soit beaucoup plus riche, complète, et fréquente. Je pense que les statistiques publiques sont vraiment fondamentales pour éclairer les politiques et c'est quelque chose qui manque cruellement dans les pays en voie de développement.

Le 3IE a un rôle très intéressant : ils financent des évaluations, ils font des revues de littératures et ils font du conseil méthodologique. Sur ce dernier point, ils sont très bons : ils montrent comment faire quantitatif et comment le coupler avec du qualitatif. Ces conseils méthodologiques permettent de recomposer des logiques en montrant qu'il est possible de faire de l'évaluation beaucoup plus modeste : si on s'aperçoit qu'on rencontre des problèmes dans les étapes intermédiaires, il ne sert à rien de dépenser des sommes considérables pour vérifier si les bébé vont mieux grâce à un programme de nutrition si on s'aperçoit que celui-ci est mal conçu, qu'il ne touche pas les personnes qui en ont besoin ou qu'il ne fournit pas une aide adaptée. C'est pourquoi on a aussi besoin d'approches plus opérationnelles : à l'AFD, l'évaluation d'impact représente d'ailleurs moins de 20% du budget de

l'évaluation. La grosse majorité des évaluations sont dites « décentralisées » : pour chaque projet qui se termine, on lance une évaluation en partenariat avec les agences et les partenaires (ministères ou autres) des pays en question, de préférence avec des évaluateurs locaux. Le référentiel d'évaluation est celui du CAD (le Comité d'Aide au Développement) qui, avant l'arrivée des RCTs, était LE référentiel de l'évaluation dans le domaine du développement. Il s'agit là de l'évaluation des processus intermédiaires où le critère d'impact ne correspond pas du tout à ce qu'était le critère d'impact dans le cas des RCTs. Il existe à ce titre une réelle confusion conceptuelle dans le monde du développement sur ce qu'on appelle l'impact.

S'il existe véritablement un besoin de faire plus d'évaluation et de la lier avec les opérations, je ne suis pas sûr qu'il faille faire de l'évaluation d'impact pour cela. Les gens du 3IE sont très ouverts dans leur conseils méthodologiques mais ce qu'ils financent ou ce qu'ils prennent en compte quand ils font les méta-évaluations, reste essentiellement des RCTs ou en tout cas du « quantitatif robuste + + ». Si l'on souhaite faire du quasi-expérimental, il faut vraiment avoir démontré par A + B qu'il était impossible de faire de l'expérimental, et le niveau d'exigence en terme de puissance statistique est très important. Cependant, ce qui conditionne la réussite d'un projet et son efficacité ne réside pas dans le fait que l'offre de moustiquaires marchent ou ne marchent pas mais dans la manière dont la gratuité est faite, par exemple, dans les circuits qui existent autour, dans la qualité du service qui elle est essentielle. Faire des évaluations d'impact pour évaluer ces choses-là c'est un peu comme écraser des mouches avec des marteaux... Des moyens surdimensionnés sont développés par rapport à ce qu'on cherche, ce qui fait que souvent on rate. Il existe des méthodes beaucoup plus modestes, basées sur de l'avant/après, des évaluations plus opérationnelles, liant quantitatif et qualitatif mais pour lesquelles on ne va pas jusqu'à de la démonstration avec contrefactuel. Bien qu'on aura toujours besoin d'évaluations d'impact, le gros de l'évaluation nécessaire aujourd'hui n'en est pas.

Exemple d'un papier écrit sur le Kenya, à propos d'une évaluation qui a montré un impact très positif de la contractualisation des professeurs sur l'éducation des enfants, mais qui n'a plus marché dès que cette politique a été généralisée : l'enjeu fondamental selon M. Bédécarrats n'est donc pas de déployer des moyens aussi considérable pour mesurer très précisément l'impact local de ce dispositif, mais de comprendre ce qui se joue dans la gouvernance de l'éducation du niveau micro, au meso et au macro.

Il y a de nombreuses recherches qui portent sur la gouvernance de la population et l'anthropologie de l'éducation, à moindre frais par rapport aux RCTs. Il y a donc une réelle question sur l'utilité et la portée des RCTs.

Les RCTs sont-ils vraiment plus cher que les méthodes qualitatives ?

Oui.

Au J-PAL pourtant, on nous a expliqué que pour un même niveau de validité interne, le même prix de l'expérience était approximativement le même.

Non. Après il faut voir ce qu'ils appellent « qualitatif ». Quand on parle de qualitatif c'est toujours un peu tarte à la crème : il faut regarder de quoi est-ce qu'on parle. Par ailleurs, parler de « même niveau de validité interne » ne signifie pas grand-chose car les critères ne sont pas les mêmes. Chez nous, une évaluation qualitative coûte au maximum 80 000 euros, tandis qu'une évaluation d'impact expérimentale peut coûter jusqu'à 600 000 ou 800 000 euros. Encore une fois, on compare des choses qui ne sont pas comparables. Collecter de la donnée c'est extrêmement coûteux. Les enquêtes terrains des instituts nationaux de statistique sont également très chères. Ce n'est pas pour autant qu'on arrêtera d'en faire car on en a besoin. La question est qu'une évaluation d'impact ne répond qu'à la question « est ce que mon microprogramme » a fonctionné ou pas. Une évaluation nationale, type INSEE, coûte sensiblement plus cher mais permet d'apporter de l'information sur quatre cents études.