Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

LES MÉTHODES MIXTES POUR L'ÉVALUATION D'IMPACT

Entretien

Adam BAÏZ

Cour des comptes, Secrétariat général

Socio-économie du travail (SET) Actuellement à la Cours des comptes où vous coordonnez la montée en puissance de l'évaluation des politiques publiques (EPP) et des partenariats académiques, vous étiez jusqu'à 2022 en poste à France Stratégie où vous avez réalisé plusieurs travaux réflexifs sur l'EPP…

Adam Baïz (AB) En effet, j'ai eu l'occasion de contribuer à divers travaux d'évaluation, mais aussi à des travaux plus réflexifs s'intéressant à la production des évaluations en France et dans le monde, à leur utilisation effective dans le débat public[1] et, plus récemment, à l'articulation des méthodes quantitative et qualitative en EPP.

SET Ce dernier aspect a donné lieu à la publication d'un guide[2] co-écrit avec Anne Revillard, chercheuse en sociologie et directrice du Liepp à Sciences Po.

AB Oui, ce guide visait à contribuer à un débat qui va au-delà de l'EPP mais qui, avec le développement des *evidence-based policies*, apparaissait central en évaluation. Historiquement, on a souvent opposé les approches quantitatives d'un côté (économétrie, analyse socio-économique, micro-simulation, etc.), et les approches qualitatives de l'autre (études de cas, observations de terrain, entretiens, etc.), et au milieu une controverse quant à leur portée et leur robustesse. Principale critique adressée aux méthodes qualitatives : des résultats non généralisables en raison d'échantillons très restreints et donc non représentatifs. Principale critique à l'encontre du quantitatif : des méthodes inadaptées pour évaluer des interventions complexes, ou reposant sur des hypothèses mathématiques irréalistes.

SET D'où une réflexion sur les spécificités et les plus-values véritables de ces approches.

AB Dans ce guide, avant de faire un panorama des méthodes pour l'évaluation d'impact, on propose une analyse des principales différences entre les approches quantitatives et les approches qualitatives. En première instance, on serait tenté de dire que les approches quantitatives et qualitatives se distinguent simplement sur le plan formel, les premières reposant sur des chiffres et les secondes sur des mots. Mais est-ce si simple ? Non, car en pratique, on retrouve des mots et des chiffres tant dans les évaluations dites quantitatives que dans les évaluations dites qualitatives. Les différences sont ailleurs, et elles sont plutôt de nature épistémologique.

[1] Baïz A., 2022, Quelles évaluations des politiques publiques pour quelles utilisations ?, Rapport, France Stratégie.
[2] Baïz A., Revillard A., 2022, Comment articuler les méthodes qualitatives et quantitatives pour évaluer l'impact des politiques publiques – Un guide à l'usage des décideurs et des praticiens, France Stratégie.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

SET Quelles sont-elles ?

AB Il y a trois principales différences relevées dans la littérature académique : les approches quantitatives visent avant tout à estimer l'amplitude des impacts là où les approches qualitatives visent avant tout à les expliquer ; les premières traitent un nombre limité de variables sur un grand nombre de cas (pour garantir la représentativité des observations), là où les secondes visent à produire une connaissance riche sur un nombre limité de cas (pour capter la complexité des phénomènes observés) ; les premières raisonnent le plus souvent de façon hypothético-déductive là où les secondes privilégient le raisonnement inductif. Mais attention à nouveau, il s'agit de différences idéal-typiques, de deux « cultures » pour citer Mahoney et Goertz (2006)[3]. En pratique, il y a déjà beaucoup de recoupements entre ces deux cultures.

SET La première différence que vous évoquez concerne la manière d'envisager la causalité…

AB Oui et on peut l'illustrer simplement. Imaginons que le taux de chômage baisse de 3 % suite à une réforme des cotisations sociales. On apprend assez vite en évaluation qu'un tel constat ne suffit pas à connaître l'impact de l'intervention, car l'évolution du chômage pourrait être due à plein d'autres facteurs (le contexte international, le prix des énergies, une politique éducative, l'implantation d'une grande entreprise, etc.). Pour connaître l'impact de l'intervention, on a besoin de savoir ce qui se serait passer sans elle, toutes choses étant égales par ailleurs. C'est le raisonnement contrefactuel. Dans la culture quantitative, la causalité contrefactuelle consiste comme en médecine à identifier un groupe avec la politique – le groupe de traitement – et un groupe identique mais sans la politique – le groupe de contrôle. Disons, « dans la région A, où la réforme s'est appliquée, le taux de chômage a baissé de 3 % tandis que dans la région B, identique à la A mais où la réforme ne s'est pas appliquée, le taux de chômage n'a baissé que de 2 % ». Par différence, le raisonnement contrefactuel nous dit que l'impact de la réforme correspond à une baisse de 1 % du taux de chômage.

Si ce raisonnement semble logique, il pose en réalité plein de difficultés en pratique. D'abord, il est impossible de trouver deux groupes parfaitement identiques en tout point. Au mieux, on peut identifier deux groupes qui se ressemblent en moyenne, sans pouvoir annuler le risque de passer à côté d'une différence subtile ou inobservable entre les deux groupes. On pourrait par exemple ne pas s'apercevoir que la région A est plus touristique que la région B et donc avoir intrinsèquement plus de facilité à créer des emplois. En ne s'en apercevant pas, l'évaluateur pourrait prêter à la réforme des effets qu'elle n'a pas en réalité. Et même si les deux groupes étaient parfaitement identiques au départ, des différences tout aussi subtiles ou inobservables pourraient apparaître au fil du temps. Par exemple, l'issue d'une élection locale pourrait finir par différencier les deux régions et leurs dynamiques de l'emploi sans que l'évaluateur ne puisse s'en apercevoir. Quand on applique une politique au groupe de traitement, le groupe de contrôle peut aussi se retrouver directement ou indirectement contaminé par la politique, si bien qu'il n'est plus vraiment un groupe de contrôle. Les entreprises concernées par la réforme pourraient par exemple produire plus d'emplois, gagner des parts de marché et ainsi faire concurrence aux entreprises de la région de contrôle. Et ainsi de suite.

SET Et la culture qualitative en matière de causalité, ce serait quoi ?

[3] Mahoney J., Goertz G., 2006, "A tale of two cultures: Contrasting quantitative and qualitative research", Political analysis, 14(3), 227-249.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

AB Dans la culture qualitative, la causalité est appréhendée de façon plutôt processuelle : on s'attache plutôt à décrire la mécanique de la causalité, à expliquer pourquoi elle génère tels ou tels impacts, à défaut de pouvoir les quantifier. Pour cela, on cherche à ouvrir la boîte noire de la politique pour analyser ses rouages. On cherche aussi à étudier l'environnement dans lequel la politique est déployée, pour relever les facteurs qui en conditionnent le succès ou l'échec. Pour revenir à notre exemple, on pourra par exemple analyser la réforme sur le plan juridique, et relever une éventuelle incohérence entre deux textes d'application. On pourra observer sur le terrain l'accueil de la réforme par les organisations syndicales, et l'éventuelle modification du rapport de forces avec les employeurs. On pourra aussi interroger ces derniers sur leurs motivations à créer de l'emploi eu égard à la réforme, et peut-être s'apercevoir d'une mauvaise compréhension de la réforme, etc. Autant d'observations et d'analyses pouvant aider à révéler des rouages clés de la politique, et des facteurs clés de son environnement d'application.

SET Mais on voit assez vite la limite du déclaratif…

AB Tout à fait, et plus généralement, ce type d'analyses ne permet pas de conclure que la réforme a bel et bien un impact, et permet encore moins de le quantifier. Un chef d'entreprise pourrait par exemple déclarer avoir créé 1000 emplois grâce à la réforme, mais peut-être en aurait-il créé aussi voire tout autant en l'absence de la réforme, grâce à un contexte international favorable. Pour estimer l'amplitude d'un impact, on a résolument besoin d'un raisonnement contrefactuel. On voit bien ainsi que les deux approches se complètent sur ce plan : les méthodes quantitatives pour estimer les impacts, et les méthodes qualitatives pour les expliquer.

SET On a donc une première différence : causalité contrefactuelle pour la culture quantitative, et processuelle pour la culture qualitative. Quelle deuxième différence épistémologique faites-vous ?

AB Elle concerne les données mobilisées : riches mais souvent peu représentatives à travers les méthodes qualitatives, et l'inverse à travers les méthodes quantitatives. Prenons l'exemple d'une évaluation portant sur les salaires. Dans une approche quantitative, on relèvera en principe une poignée de données (appelées variables), par exemple le salaire moyen et le chiffre d'affaires de l'entreprise, et on le fera sur un échantillon suffisamment large et astucieusement constitué pour qu'il soit représentatif de l'ensemble des entreprises. Dans une approche qualitative, on cherchera plutôt à comprendre les mécanismes sous-jacents à la fixation des salaires. On interrogera chaque entreprise sous toutes ses coutures (son tissu économique, son histoire syndicale, ses enjeux de trésorerie, etc.) pour en révéler les spécificités, les nuances, la complexité. Mais comme ces investigations sont coûteuses, on le fera en principe sur un petit nombre d'entreprises. Si bien que les observations réalisées ne seront pas représentatives et pourront difficilement être généralisées à l'ensemble de la population des entreprises.

SET Pouvez-vous en dire plus sur cet enjeu de représentativité ?

AB Comme il est souvent coûteux, voire impossible, d'observer toute la population qui nous intéresse dans le cadre d'une étude, on se contente en pratique d'observer un échantillon. Tout l'enjeu est d'avoir un échantillon représentatif de la population, c'est-à-dire un échantillon qui ressemble en tout point à la population dont il est issu. C'est la condition pour pouvoir généraliser à l'ensemble de la population les observations faites sur l'échantillon. Par exemple, si on s'intéresse à l'ensemble des entreprises d'un pays, on se concentrera en pratique sur un échantillon de ces dernières en

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

veillant à y retrouver les mêmes caractéristiques en termes de secteurs économiques, de tailles, de statuts juridiques, etc.

La difficulté, bien sûr, c'est qu'il est impossible d'avoir un échantillon ressemblant en tout point à la population. Pour reprendre notre exemple, les entreprises de l'échantillon ne pourront jamais totalement reconstituer toutes les caractéristiques des entreprises qui n'en font pas partie. Et pour cause, chaque entreprise a forcément une caractéristique – ou une combinaison de caractéristiques – qui la rend unique. Ainsi, dans la culture quantitative, on fait l'effort de constituer des échantillons larges – souvent quelques centaines à quelques milliers d'unités – et astucieusement sélectionnés – en les tirant au sort par exemple, pour éviter de sur-représenter certaines caractéristiques – afin d'avoir le maximum de similarités avec la population-mère. Tout en faisant cela, le statisticien prendra soin d'exprimer les résultats sous la forme d'intervalles de confiance : si l'échantillon par exemple révèle un salaire moyen de 2300 €, on dira par exemple que le salaire moyen dans l'ensemble de la population est estimé à 2300 € ± 100 € et avec une incertitude de 5 % (c'est-à-dire qu'il y a 5 % de chances, eu égard à l'aléas d'échantillonnage, pour que le véritable salaire moyen dans la population globale soit en réalité plus petit que 2200 € ou plus grand que 2400 €).

Autant dire que les échantillons mobilisés dans les méthodes qualitatives – souvent de quelques unités seulement – sont en principe très peu représentatifs, et leurs résultats difficilement généralisables à l'ensemble de la population. Ceci dit, en s'attachant à comprendre la complexité de certains cas particuliers, ces méthodes qualitatives peuvent être particulièrement efficaces pour prédire un phénomène précis. Par exemple, elles peuvent aider à prédire l'évolution des salaires à la suite d'une élection syndicale dans une entreprise donnée, si on a bien su expliquer les liens entre le fonctionnement syndical et la fixation des salaires. En un sens, les méthodes quantitatives permettent de généraliser des moyennes à une population globale, là où les méthodes qualitatives autorisent des extrapolations sur des situations précises, ce qui peut être très utile aussi.

SET Et pour finir, quelle troisième différence épistémologique feriez-vous ?

AB Une différence en termes de raisonnement avec plutôt de l'hypothético-déductif dans la culture quantitative et de l'inductif pour la culture qualitative. Dans le premier cas, on commence avec une hypothèse et on développe un protocole pour la confirmer ou l'infirmer. Par exemple : « Hypothèse : les maladies professionnelles en entreprise augmentent avec l'ancienneté ». Là on pourrait conduire un test de corrélation : on fait plein d'observations et on regarde le nuage de points pour voir si on a une corrélation positive. Dans la culture qualitative, le raisonnement est plutôt inductif. L'idée c'est qu'au départ, il y a forcément des hypothèses dans l'esprit du chercheur, et puis ce dernier part sur le terrain à la recherche de cas dits déviants. Par exemple, il pourrait tomber sur le cas de « Albert, cadre, 20 ans d'ancienneté et aucune maladie professionnelle, contrairement à Léo, ouvrier, et 3 ans d'ancienneté ». Là, le chercheur se dit : « ah tiens, en fait, il n'y a pas que l'ancienneté qui joue dans les maladies professionnelles mais peut-être aussi le statut de l'employé, ses conditions de travail, etc. » donc il révise son hypothèse et repart sur le terrain à la recherche de nouveaux cas déviants. Dans la culture qualitative, il y a ainsi un aller-retour beaucoup plus assumé entre le terrain et la réflexion, et une fois que l'hypothèse ne rencontre plus – ou quasiment plus – de cas déviants, on dit qu'elle est saturée et ça devient une théorie.

Dans notre guide, on développe aussi un point sur le traitement de la subjectivité dans les méthodes qualitatives en notant que dans la culture qualitative, les évaluateurs sont plus enclins à interroger la subjectivité des acteurs, leurs représentations, leurs

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

perceptions… mais aussi plus enclins à interroger – de façon réflexive – leur propre subjectivité en se demandant par exemple « j'utilise tel indicateur, mais est-ce bien pertinent ? », « est-il possible que je sois soumis à un biais de confirmation ? », « comment mettre en difficulté mes préconceptions et produire des résultats plus rigoureux ? », etc.

SET Est-ce qu'on ne voit pas là déjà s'esquisser une forme de complémentarité entre les cultures ?

AB Si en effet, les méthodes quantitatives visent à estimer les impacts, là où les méthodes qualitatives visent à les expliquer, les premières à produire des résultats significatifs sur une poignée de variables, et les secondes à développer l'analyse de cas particuliers, etc. En outre, les approches quantitatives peuvent magistralement se tromper, y compris avec beaucoup de données, des formules mathématiques sophistiquées, et de la puissance de calcul. On peut citer au moins deux exemples. Le premier, c'est le paradoxe de Simpson, qui traduit notamment l'idée qu'un nuage d'observations peut cacher des sous-nuages aux allures bien différentes. Par exemple, un certain nuage d'observations, représentant 10 000 salariés avec leur salaire en abscisse et leur productivité en ordonnée, pourrait suggérer une corrélation négative entre ces deux variables. Le statisticien pourrait se satisfaire de ce résultat obtenu en considérant qu'il a été calculé à partir d'un large échantillon représentatif de la population des salariés. Mais il pourrait magistralement se tromper si ce nuage cache en réalité plusieurs sous-nuages correspondant par exemple à des catégories socio-professionnelles bien particulières, et faisant apparaître autant de corrélations à l'inverse positives (voir le graphique FIG. 1). Avant de calculer des corrélations, des moyennes ou encore des ratios, il est crucial de bien choisir les catégories sur lesquelles ces calculs seront réalisés : les méthodes qualitatives, en étant particulièrement attentives aux spécificités du terrain, sont très utiles de ce point de vue.

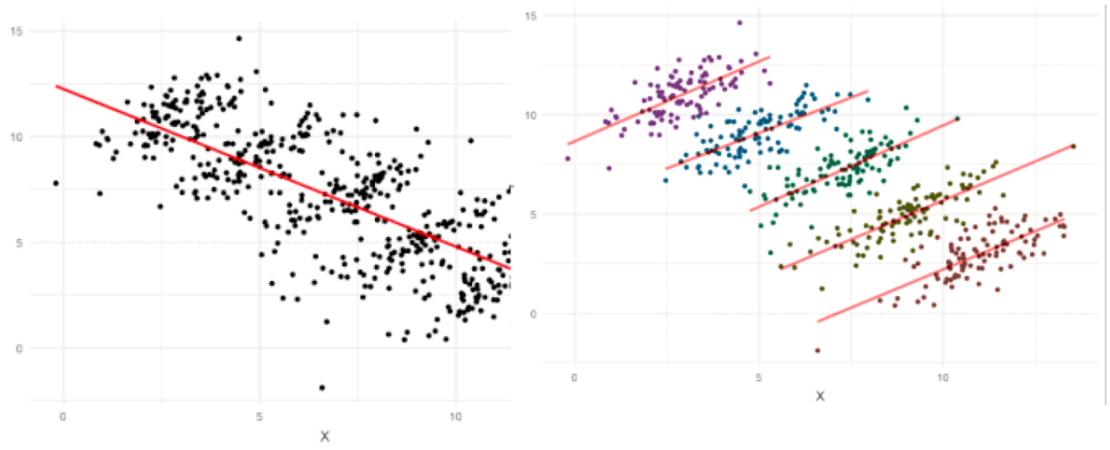

FIG. 1 – Illustration du paradoxe de Simpson

Source : Wikipedia https://en.wikipedia.org/wiki/Simpson%27s_paradox (consulté le 01/11/2024).

Lecture : Le nuage de points, à gauche, laisse apparaître une corrélation négative entre les deux variables x (abscisse) et y (ordonnée). En réalité, ce nuage de points pourrait cacher des « sous-nuages » bien spécifiques (le nuage violet correspondant par exemple aux ouvriers, le nuage bleu aux artisans, etc.) au sein de chacun desquels la corrélation est désormais positive. Et chacun de ces « sous-nuages » pourraient en cacher d'autres…

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

SET Les statisticiens ont quand même l'habitude d'introduire des variables de contrôle type âge, sexe, origine géographique, etc.

AB Certes mais pour un sujet d'étude donné, les bonnes variables de contrôle, les bonnes catégories, sont peut-être ailleurs. Pour les identifier, il est essentiel d'aller sur le terrain pour mieux en comprendre les spécificités et la complexité.

SET Et le second exemple ?

AB Le second exemple est le *p-hacking* : cela consiste à passer en revue des variables par dizaines sur de larges échantillons jusqu'à tomber sur une forte *p-value* suggérant artificiellement un résultat significatif (*i.e.* robuste). Donnons un exemple. On pourrait considérer un million d'individus avec des dizaines de pages de données sur chacun d'eux, et rechercher deux caractéristiques que partagent certains individus à l'exclusion de tous les autres. Par exemple, sur ce million d'individus, on pourrait découvrir (par accident ou à force de chercher !) que 1000 d'entre eux sont nés un 22 février entre 12 h et 14 h et sont de grands fans de Tintin. Par chance, on pourrait vérifier qu'aucun des autres individus – ou presqu'aucun – ne partage ni l'une ni l'autre de ces deux caractéristiques. Statistiquement, on pourrait alors établir un lien de corrélation très significatif (c'est-à-dire une très faible *p-value*) entre ces deux caractéristiques. C'est le genre de démarche qui prévaut dans des études aux résultats sensationnels : « Si votre enfant naît un 22 février entre 12 h et 14 h, il aura de très grandes chances d'être fan de Tintin ! ». La corrélation sera robuste en effet, mais sans aucun fondement théorique. À nouveau, les méthodes quantitatives seules peuvent se tromper ou être trompeuses.

SET Mentionner ces risques, c'est déjà suggérer l'intérêt des méthodes mixtes…

AB Oui, les méthodes mixtes visent à tirer profit de la complémentarité entre les approches quantitatives et les approches qualitatives. La littérature fait état de 8 formes de complémentarité (Aguilera et al., 2021)[4] : la confirmation, l'enrichissement, la complexification, la systématisation, la généralisation, la triangulation, la complémentarité, le développement. Vis-à-vis d'une approche quantitative, une approche qualitative peut servir à augmenter la portée d'une théorie avec un nouvel exemple (confirmation), à illustrer la théorie en approfondissant un cas particulier (enrichissement), à réviser la théorie en lui trouvant un contre-exemple (complexification). Vis-à-vis d'une approche qualitative, une approche quantitative peut servir à renforcer la significativité des résultats, grâce à de plus grands échantillons (systématisation) ou à extrapoler à de nouvelles situations (généralisation). Enfin, conjuguer les deux types d'approches peut permettre de corroborer (ou non) un résultat à partir d'un faisceau d'indices (triangulation), de répondre à des questions connexes (complémentarité) ou de s'enrichir à partir d'un prisme méthodologique ou disciplinaire différent (développement).

SET Peut-on citer des exemples précis d'application de méthodes mixtes ?

AB Prenons l'Analyse qualitative comparée (AQC). Cette méthode s'intéresse aux facteurs de succès ou d'échec d'une politique. La méthode consiste à lister les facteurs potentiels de succès ou d'échec, à observer la politique au croisement de ces facteurs (présents ou non, faibles ou forts) et à repérer les facteurs clés et les effets « cocktail ».

[4] Aguilera, T., & Chevalier, T. (2021). Les méthodes mixtes pour la science politique. Revue française de science politique, 71(3), 365-389.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

Par exemple, dans une étude de 2013 évaluant une politique de lutte contre les grossesses adolescentes à Londres (Blackman, 2013)[5], Blackman cherche à comprendre pourquoi la politique est un succès dans certains quartiers et un échec dans d'autres quartiers. Grâce à quelques investigations préliminaires, Blackman identifie un certain nombre de facteurs pouvant potentiellement favoriser le succès ou l'échec de la politique : la proportion de jeunes de moins de 18 ans, la diversité ethnique, la prévalence de la consommation de drogue, etc. En comparant ensuite, et de manière systématique, des quartiers où ces facteurs sont présents ou absents, Blackman s'aperçoit que la politique est un succès dans tous les quartiers où la diversité ethnique est forte : cette diversité ethnique apparaît comme un facteur suffisant de succès de la politique. Il s'en étonne, et s'aperçoit qu'une forte diversité ethnique cache en réalité un tissu associatif particulièrement actif dans la sensibilisation aux infections sexuellement transmissibles. Grâce à ses travaux, on tient alors une recommandation concrète pour augmenter les chances de succès de la politique dans tous les quartiers : renforcer ce tissu associatif dans les quartiers. Avec cette méthode, on peut aussi identifier les facteurs nécessaires au succès de la politique, aux facteurs aidants, ou encore aux facteurs inutiles ou contre-productifs.

SET Comme pour une Analyse de contribution ?

AB Alors, dans une analyse de contribution, on va plutôt chercher à identifier les composantes d'une politique qui contribuent le plus aux impacts identifiés. Une politique, c'est rarement un seul levier, c'est souvent un ensemble d'actions qui sont déployées conjointement. Il s'agit donc de décomposer la politique en ses diverses composantes, proposer des récits possibles de contribution, et tester ces hypothèses sur le terrain (entretiens, observations, etc.) pour repérer les contributions clés. Prenons un exemple d'évaluation portant sur la filière bois au Congo (Delahais et al., 2017)[6]. Plusieurs actions ont été déployées pour y promouvoir des pratiques écoresponsables : proposer de nouveaux indicateurs, produire des données ou encore identifier des solutions innovantes. On a observé au fil des années des pratiques de plus en plus écoresponsables dans la filière bois. Laquelle de ces actions y a le plus contribué ? Pour répondre à cette question, on peut comparer les différents sites d'exploitation forestière où telle ou telle action est plus ou moins prégnante, interroger les acteurs concernés, ou encore établir la chronologie des événements. Les évaluateurs ont en l'occurrence découvert que les acteurs de la filière bois ont changé leurs pratiques à la sortie d'un rapport proposant de nouveaux indicateurs, et ont reconnu avoir modifié leurs pratiques après s'être appropriés ces nouveaux indicateurs. Cette approche vise ainsi à identifier les actions qui sont nécessaires ou suffisantes pour avoir tel impact et, dans une logique d'efficience, de se débarrasser des actions redondantes, inutiles, voire contre-productives. En définitive, dans l'Analyse qualitative comparée, on s'intéresse aux facteurs contextuels de succès ou d'échec de la politique, tandis que dans l'analyse de contribution, on s'intéresse surtout aux composantes même de la politique qu'il est important de mettre en œuvre, de retirer ou d'ajuster.

SET Ces approches font jouer un rôle essentiel à une théorie de l'intervention (ou théorie de l'action) préalablement définie. Dans votre thèse de doctorat en sciences

[5] Blackman T., Wistow J., Byrne D., 2013, "Using qualitative comparative analysis to understand complex policy problems", Evaluation, 19(2), 126-140.
[6] Delahais T., Toulemonde J., 2017, "Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management?", Evaluation, 23(4), 370-388.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

de gestion (Baïz, 2018)[7], vous développez une méthode systématique pour baliser ce type d'exercice…

AB J'ai en effet cherché à développer une méthode pour apprécier systématiquement l'impact d'une politique au regard de la faisabilité de sa mise en œuvre. J'ai simplement appelé cette méthode « l'évaluation faisabilité-impact ». En deux mots, cette méthode consiste à lister tout ce que les acteurs sont censés faire, et toutes les influences qu'ils sont censés avoir les uns vis-à-vis des autres. Par exemple, untel est censé payer tel somme d'argent en cas de pollution, untel est censé encourager tel autre à acheter une voiture électrique, etc. Puis, on vérifie si chaque chose a lieu sur le terrain comme elle était prévue sur le papier (la loi, le règlement, etc.). Cette vérification systématique permet d'identifier des biais et obstacles de mise en œuvre pouvant altérer son impact. Prenons l'exemple d'une taxe carbone. Plutôt que de l'envisager comme une simple somme d'argent, on pourrait lui associer cette logique d'action : « si une entreprise pollue, alors elle paye une somme d'argent proportionnelle à ses émissions ce qui la décourage de polluer à nouveau ». À partir de là, si l'on constate que l'impact observé n'est pas celui recherché, on va pouvoir suspecter deux obstacles : soit que la somme payée n'est pas assez dissuasive (un écueil auquel on pense intuitivement), soit que les entreprises polluantes n'ont pas payé leur dû (un écueil auquel on pense moins !), soit les deux.

SET En fait, vous systématisez la notion de diagramme logique des impacts…

AB Tout à fait. Je propose un langage – à la fois graphique et algorithmique – pour décrire chaque politique de façon plus ou moins granulaire. Par exemple, pour la taxe carbone, on pourra préciser que tel acteur doit aller sur tel logiciel pour payer sa taxe. En relevant une action aussi précise que celle-ci, cela nous obligera en tant qu'évaluateur à vérifier qu'elle a bien eu lieu, et peut-être alors s'apercevoir que c'est un *bug* informatique qui a enrayé toute la politique et altéré son impact. Ce langage permet aussi d'articuler des politiques entre elles, et d'étudier leurs interconnexions et interdépendances. Par exemple, de relever que la taxe carbone peut à la fois jouer un rôle dissuasif sur la consommation de carburants, mais aussi un rôle pour le financement des énergies vertes. Enfin, ce langage vise à se défaire de certains automatismes quand on conçoit ou révise des politiques. On pourrait par exemple considérer une taxe « non-financière » du type : « si une entreprise pollue, alors elle rend public le total de ses émissions ce qui la décourage de polluer à nouveau ». La logique d'action est la même – « autoriser une action X mais imposer une action dissuasive Y derrière » – mais selon des modalités concrètes différentes.

SET Vous envisagez donc l'évaluation de faisabilité-impact, selon vos termes, comme un outil permettant d'articuler cultures qualitative et quantitative en matière d'évaluation.

AB Oui, c'est une méthode mixte qui cherche à concilier ces deux cultures : il y a d'un côté ce souci de révéler et de comprendre les mécaniques causales à l'œuvre, et de l'autre le souci de quantifier les phénomènes observés. La méthode prévoit notamment des calculs en probabilités : par exemple, un acteur A a encouragé un acteur B à faire une

[7] Baïz A., 2018, De l'innovation des instruments de politique publique : développement d'une méthode de conception combinatoire autour d'un langage algorithmique et application au dispositif des certificats d'économie d'énergie (thèse de doctorat, Université Paris sciences et lettres).

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°15, 2024-1. DOI : https://doi.org/10.48611/isbn.978-2-406-18609-0.p.0067.

action X, si la probabilité que l'acteur B fasse l'action X a augmenté suite à l'intervention de l'acteur A, et si cette probabilité n'aurait pas augmenté autrement.

L'articulation des méthodes quantitatives et des méthodes qualitatives posent plusieurs défis. Il y a d'abord le défi du dialogue entre des méthodes dont la technicité est croissante, par exemple lorsqu'il s'agit d'articuler des approches processuelle et contrefactuelle de la causalité. Il y a le défi aussi de constituer des équipes suffisamment multidisciplinaires, et de faire travailler ensemble des experts aux profils variés. En termes de ressources, mobiliser et croiser plusieurs méthodes peut être plus coûteux. Enfin, il y a toujours le risque d'un usage cosmétique des méthodes mixtes : quelques statistiques par-ci, et quelques entretiens par-là. Mais l'effort en vaut la chandelle : les méthodes mixtes enrichissent la connaissance et éclairent plus utilement encore l'action publique !