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.0081.

LES DOCTRINES INSTITUTIONNELLES EN MATIÈRE D'ÉVALUATION DE L'IMPACT DES POLITIQUES PUBLIQUES : UN COMMENTAIRE

Solveig GRIMAULT

IRES

Ces dernières années, l'évaluation des politiques publiques (EPP) et les différentes approches selon lesquelles elle peut être conduite ont fait l'objet d'une série de publications de France Stratégie[1] (Bono *et al.*, 2018 ; Desplatz *et al.*, 2019 ; Desplatz et al., 2016 ; Baïz et al., 2022 ; Baïz, 2022) et de deux études annuelles du Conseil d'État (Conseil d'État, 2019 ; Conseil d'État, 2020). Outre sa contribution à l'EPP, à travers les comités d'évaluation qu'elle coordonne ou dont elle assure le secrétariat, France Stratégie s'est ainsi engagée dans la production de « travaux réflexifs sur l'exercice même de l'évaluation, ses méthodes et ses pratiques » (De Margerie, 2022). Le Conseil d'État indique quant à lui souhaiter « concrétiser une réflexion latente depuis plusieurs années [… et] mettre à profit son point de vue privilégié sur la conduite des politiques publiques en proposant [une sorte] d'évaluation de l'évaluation » (Conseil d'État, 2020, p. 10) confirmant une intention déjà présente dans son étude de 2019 consacrée aux expérimentations[2]. Dans le prolongement de l'entretien avec Adam Baïz reproduit *supra*, l'objet de ce commentaire est de discuter les éléments de doctrine figurant dans l'ensemble de ces travaux institutionnels, notamment concernant l'articulation des approches quali- et quantitative en EPP. Les méthodes quantitatives, appréciées pour leur appareillage formel, mesurent l'impact, entendu statistiquement comme effets moyens, d'une intervention en lui associant un contrefactuel. Les méthodes qualitatives décrivent les mécanismes et les facteurs d'échec ou de succès de l'intervention à partir d'entretiens, d'observations directes et d'études de cas. Les secondes sont trop souvent reléguées à un rôle subalterne, réduit à l'interprétation de l'impact quantitatif voire à l'illustration. Leur portée est pourtant bien plus grande pour autant qu'on en saisisse pleinement les modalités. Le guide de Baïz et Revillard (2022) constitue la publication institutionnelle allant le plus loin en ce sens. Le présent commentaire en souligne les avancées tout en pointant un certain nombre de simplifications et de raccourcis, notamment dans la manière d'envisager le « cas » en tant qu'instance d'observation qualitative d'une intervention. Le commentaire aborde les ambiguïtés du partage entre qualitatif et quantitatif, rappelle le potentiel de l'étude qualitative en matière d'inférence causale et, se référant à la tradition pragmatiste, indique en quoi les approches qualitatives permettent d'élargir la façon d'assigner une valeur à une intervention, en particulier dans une direction moins technocratique.

[1] Institution autonome placée auprès du Premier ministre.
[2] Alors que celles-ci avaient trouvé un nouveau fondement dans la Constitution avec la loi du 28 mars 2003.

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.0081.

I. LA PRÉDOMINANCE DES MÉTHODES QUANTITATIVES DANS LE CHAMP DE L'ÉVALUATION D'IMPACT

Comment s'est installé le « climat » actuel du champ de l'évaluation d'impact et le rôle subalterne assigné aux travaux qualitatifs ?

I.1. LES ÉTUDES FRANCE STRATÉGIE

Les premiers travaux publiés par France Stratégie en 2016 s'ouvrent sur l'idée qu'une évaluation d'impact, définie simplement, vise à établir la situation qu'aurait connue la société en l'absence de la politique évaluée (Desplatz et Ferracci, 2016). La reconstitution d'une situation fictive, dite contrefactuelle, y est ainsi présentée comme la clé d'une évaluation d'impact réussie. Le guide correspondant, conçu à l'usage des décideurs et praticiens, présente les principales méthodes permettant de mesurer l'impact d'une politique sur ses bénéficiaires directs (p. 5-24). Dans cet ensemble, les auteurs indiquent que « l'expérimentation contrôlée, ou aléatoire, constitue la méthode privilégiée par de nombreux évaluateurs, car elle présente l'avantage de construire la situation contrefactuelle de façon à la fois rigoureuse et intuitive ». Pour autant, il n'est pas nécessairement aisé de saisir ce que le terme « d'intuitif » cherche précisément à connoter : la simplicité des moyens techniques, des hypothèses de base, la lisibilité des résultats pour des non spécialistes du champ statistique (Zamora, 2011) ? Philippe Zamora estime en effet que compte tenu de ces dernières caractéristiques, « l'intérêt [de la méthode d'évaluation aléatoire] est […] à la fois scientifique et politique » (*ibid.*). Après avoir rappelé le principe de la mesure de l'effet causal du traitement dans la méthode de l'expérimentation aléatoire – écart entre la moyenne de la variable de résultat au sein du groupe de traitement, et la moyenne de la même variable au sein du groupe de contrôle –, les auteurs présentent une série de développements techniques qui ne seront pas évoqués ici. D'autres méthodes mobilisables sont également citées qui, lorsque les conditions ne sont pas réunies pour mettre en œuvre une expérimentation contrôlée, permettent-elles aussi la construction d'un contrefactuel pour parvenir à établir la causalité.

Un chapitre spécifique est ensuite consacré à la question de l'explication de l'échec ou du succès des politiques soumises à évaluation (p. 29-46), où apparaissent les méthodes qualitatives. Les méthodes jusque-là recensées « [permettant] d'évaluer l'impact d'un dispositif sur un ou plusieurs critères mais pas les raisons de son échec ou de son succès », les auteurs invitent à recourir à d'autres approches susceptibles de compléter les résultats. C'est à ce titre que sont introduites les approches qualitatives – notamment – qui permettent de comprendre les facteurs d'échec et de succès de la politique évaluée. Le qualitatif prendrait ainsi en charge des dimensions qui « ne relèvent pas de l'évaluation d'impact, mais [qui] peuvent fournir des clés de compréhension des résultats portant sur les effets de l'intervention publique considérée » – la compréhension d'un taux de non-recours important à un dispositif, par exemple. Un partage des tâches est ainsi suggéré : le quantitatif évalue l'impact d'un dispositif et le mesure, mais « ne renseigne pas nécessairement sur les mécanismes par lesquels [celui-ci] agit » ; le qualitatif, lorsqu'il est mobilisé en complément, permet d'interpréter les résultats issus de l'analyse quantitative, en proposant des explications à partir d'éléments que le quantitatif ne parvient pas à saisir (croyances, conceptions, représentations sociales). Enfin, le

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.0081.

qualitatif est également invoqué sur le registre de l'illustration, par exemple de la diversité des situations ou de l'hétérogénéité des pratiques et des comportements.

Ces travaux de France Stratégie ont trouvé un débouché dans un colloque organisé en partenariat avec l'Assemblée nationale, le « Printemps de l'évaluation 2018 », qui a donné la plus grande place aux approches contrefactuelles. Deux nouveaux documents de travail s'inscriront ensuite dans cette même perspective et cette même conception de l'évaluation d'impact pour en documenter la pratique en France et à l'étranger (Bono *et al.*, 2018 ; Desplatz et Lacouette-Fougère, 2019).

I.2. LES ÉTUDES DU CONSEIL D'ETAT

C'est dans ce « climat » qu'interviennent les études annuelles du Conseil d'État. La première (Conseil d'État, 2019), relative aux expérimentations, consacre une section importante au cadre méthodologique nécessaire à la conception, la conduite et l'évaluation des expérimentations (p. 43-67). Dès lors que celles-ci recourent à un échantillon pour apprécier les effets d'une mesure sur une population cible, le rapport suggère qu'elles mobilisent dans la mesure du possible la méthode d'expérimentation aléatoire ou randomisée, « souvent présentée comme la méthode expérimentale la plus rigoureuse d'un point de vue scientifique ("gold standard")[3] ». Certains inconvénients sont néanmoins soulevés – son coût et certains biais méthodologiques. Le rapport invite par ailleurs à combiner méthodes quantitative et qualitative pour apprécier l'efficacité d'une mesure, sans préciser toutefois les modalités et apports potentiels de cette association, si ce n'est par la mention des indicateurs chiffrés et qualitatifs qu'il conviendrait alors de mobiliser – et au titre de ces derniers « notamment l'opinion des sujets de l'expérimentation (l'interrogation individuelle ou collective des acteurs, l'observation directe par l'évaluateur, la monographie) ».

La seconde étude, spécifiquement consacrée à l'EPP, dresse un large panorama des enjeux et de l'évolution des pratiques évaluatives, et caractérise dès son introduction la singularité de l'évaluation d'impact dans cet ensemble (Conseil d'État, 2020). Cette dernière est d'emblée située dans le champ des méthodes quantitatives : visant à « mesurer, à partir de méthodes statistiques et économétriques, les effets propres d'une politique publique au regard des objectifs qui lui sont assignés », l'évaluation d'impact est « une branche spécifique de l'évaluation, principalement réalisée par des chercheurs en économie » (*ibid.*, p. 26). L'étude consacre des développements détaillés aux exigences des pratiques évaluatives, concernant à la fois les données et leur traitement, les grands principes et conditions de fiabilité d'une évaluation, les rapports avec le Parlement et les parties prenantes, les méthodes mobilisables enfin et leurs limites (*ibid.*, p. 65-70). Les spécificités des méthodes quantitatives et qualitatives s'énoncent dans les mêmes termes que dans les travaux précédemment cités : les premières mesurent les impacts, à l'appui de la construction statistique d'un contrefactuel ; les secondes

[3] On peut toutefois signaler que dans des domaines de la recherche mobilisant fortement les essais contrôlés randomisés, ceux-ci constituent le second niveau de preuve, le plus haut niveau de preuve s'obtenant par les méta-analyses (selon la hiérarchie des niveaux de preuve établie par EUROPATI- European Patients'Academy on Therapeutic Innovation). Le « gold standard » est donc établi sur la base d'une information filtrée, et à partir d'une série de tests contrefactuels pour un traitement donné. Certains auteurs soulignent que les études randomisées doivent fréquemment procéder à des essais répétés pour obtenir un résultat (moyen) assuré (Concato *et al.*, 2000). Or dans le champ des politiques de l'emploi, la possibilité de répéter l'expérimentation comme un traitement – au sens de « l'expérimentation pour l'évaluation » (Duclos, 2024) – est difficilement envisageable et supposerait vraisemblablement des ressources trop conséquentes pour être véritablement mobilisable. La possibilité de retrouver des conditions comparables est faible qui plus est, dans ces matières. Signalons que dans le champ des politiques d'emploi, la méta-analyse de référence est celle de Card, Kluve et Weber, mentionnée dans le guide de France Stratégie de 2016 (Card *et al.,* 2015).

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.0081.

recensent les effets, permettent d'identifier et de comprendre les mécanismes explicatifs, voire d'« éclairer les résultats parfois contradictoires ou inattendus des analyses quantitatives, en prenant en compte le jeu des acteurs et la réalité sociale ». Le cœur du travail évaluatif reste un attribut des analyses quantitatives, que les travaux qualitatifs peuvent venir compléter : « L'approche qualitative est complémentaire des analyses quantitatives qui conduisent à formuler des hypothèses causales ». Cette conception de la distribution du travail évaluatif ne s'est pas démentie avec l'audience croissante de la méthode de l'expérimentation aléatoire contrôlée qui, du point de vue du Conseil d'État, a contribué à techniciser l'évaluation « afin de mieux identifier les relations causales entre une politique et les modifications qu'elle a pu entraîner dans l'état de la société ».

Invité à contribuer aux travaux du Conseil d'État en tant que président du Cnis, Patrice Duran[4], semble avoir naturalisé cette même partition du champ de l'évaluation d'impact, tout en développant un point de vue critique. Duran distingue de longue date l'évaluation d'impact, qui requiert selon lui des méthodes quantitatives afin de « raisonner sur les impacts et conséquences de l'action » (Duran, 2010, p. 16)[5], de « l'évaluation de processus » (Duran *et al.*, 2018, p. 5) ou de mise en œuvre. Il insiste sur le fait que « l'accent mis sur les conséquences [d'une intervention] est loin de signifier que seules comptent les conséquences indépendamment de ce qui les produit » (Duran, 2018, p. 11) : la mise en œuvre constitue un phénomène « complexe » et « central » qu'il convient d'intégrer pleinement aux logiques évaluatives (*ibid.*). Duran s'inquiète du « risque d'une identification de l'évaluation à [la] seule figure » de l'évaluation d'impact (p. 187-188) : « on comprend aisément le succès actuel de l'évaluation d'impact auprès des administrations dans la mesure où [… elle] les exonère […] d'une réflexion de fond sur leur propre fonctionnement. [Mais …] il [est particulièrement] dommageable [… de ne pas être en] capacité de lier des résultats à des logiques organisationnelles et […] professionnelles. […] la connaissance des mécanismes qui ont conditionné la mise en œuvre d'une politique est […] nécessaire à ceux qui [l'ont portée pour réfléchir à] la qualité du management comme condition de la qualité des résultats » (p. 187-188).

Le rapport du Conseil d'État revient enfin sur la question des méthodes dans une partie plus prospective relative aux moyens à développer pour « mieux évaluer » (p. 89-117). Rappelant la critique développée par Alain Supiot[6], il souligne qu'une évaluation ne saurait s'en tenir à une approche exclusivement comptable ou quantitative. Mais la contribution du qualitatif reste seconde, quoique celui-ci soit considéré comme « indispensable pour […] comprendre les mécanismes de mise en œuvre ». Il permettrait en particulier d'apprécier « les effets potentiels et les risques que peut comporter une réforme, au regard notamment de son acceptabilité par le corps social ». À cela s'ajoute l'idée que les études d'impact « gagneraient à inclure […] un regard prospectif partagé avec des acteurs non gouvernementaux, comme les partenaires sociaux, selon la méthode qui était celle du Commissariat général du plan et qui fait, aujourd'hui encore, le succès des analyses de France Stratégie[7] ».

---

[4] Par ailleurs professeur émérite des universités, École normale supérieure Paris Saclay.

[5] Dans un article paru en 2021, Duran soulignera à nouveau combien l'évaluation d'impact est marquée par sa « forte technicité au plan économétrique », situant ainsi explicitement l'évaluation d'impact dans le champ des méthodologies quantitatives (Duran, 2021).

[6] Supiot A., 2015, La gouvernance par les nombres, Fayard.

[7] Cette méthode était alors qualifiée « d'expertise concertée » par le Commissariat général du Plan (CGP), ainsi que le rappellent Olivier Baguelin et Laurent Duclos (Baguelin *et al*.,2023). Ils soulignent en particulier que le rapport Viveret, élaboré au sein du CGP et qui a posé les fondements d'une approche française de l'évaluation, était porteur de l'exigence d'un certain pluralisme en matière d'EPP. L'évaluation pluraliste signifiait alors partagée, ou concertée.

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.0081.

Parmi les vingt propositions qu'il formule, le Conseil d'État suggère, comme en 2019, de « favoriser une approche pluridisciplinaire, prenant en compte de façon équilibrée les données quantitatives et qualitatives, et s'appuyant sur la diversité des sciences sociales ». Dans un souci de rigueur, le rapport prend également le soin de pointer « quelques faiblesses persistantes » et quelques « pièges et […] limites méthodologiques » dans l'exposé des différentes méthodes mobilisables (p. 68-69). Ainsi « l'identification d'un lien causal entre une politique et des résultats peut se révéler plus complexe qu'il n'y paraît. Les interactions, la multiplicité des effets, peuvent induire en erreur ». Le « risque de confusion entre cause et corrélation » dans le cas de l'évaluation quantitative d'impact est également pointé, à l'appui d'une note du Conseil d'analyse économique (CAE, 2013)[8]. Du côté des travaux qualitatifs, un certain flou perdure entre les lignes sur la portée et la qualification précises de leur contribution aux évaluations, comme sur leurs procédés méthodologiques. De là découle peut-être leur mise à l'écart persistante des évaluations d'impacts proprement dites, si ce n'est comme auxiliaires – en complément – des travaux quantitatifs. Quelque chose chercherait encore à s'expliciter concernant les apports des uns et des autres à une démarche évaluative. Une incertitude semble demeurer à propos des travaux qualitatifs, que le guide publié en 2022 par France Stratégie va venir lever.

II. MÉTHODES ET DÉMARCHE D'ÉVALUATION

La focalisation des débats sur les *méthodes* d'évaluation, accentuée par l'opposition entre quantitatif et qualitatif, a supplanté une autre dimension, celle de la *démarche* évaluative elle-même. L'évaluation « basée sur la théorie » (*theory based evaluation*, TBE), d'inspiration pragmatiste, mise précisément sur le potentiel d'une *démarche* permettant d'organiser de façon cohérente des travaux se réclamant de méthodes diverses. Cela conduit à souligner l'importance et l'utilité de la tradition pragmatiste pour l'EPP.

II.1. DÉMARCHE ÉVALUATIVE ET PLURALISME

Par-delà les questions de méthodes, le Conseil d'État observe que « l'évaluation des politiques publiques, "démarche à caractère institutionnel et à visée opérationnelle" [Duran, 2018, p. 7], ne saurait se réduire aux procédures et aux outils mis en œuvre, ni avoir pour objectif la publication d'une étude scientifique » (Conseil d'État, 2020, p. 24-25). Les méthodes mobilisées, en d'autres termes, ne sauraient saturer l'espace de l'évaluation qui ne peut s'affranchir de la mise en œuvre d'une démarche évaluative. Pour le Conseil d'État « l'évaluation est une démarche qui vise à analyser les effets d'une politique publique pour en estimer la valeur ajoutée », une démarche marquée par « la recherche de connaissance et de compréhension » et qui emporte à ces divers titres des exigences particulières. La notion de démarche apparaît principalement en lien avec les grands principes généralement admis pour garantir la fiabilité d'une évaluation (p. 62-65). L'indépendance, tout d'abord, de l'évaluation doit ainsi être appréhendée en « [tenant] compte des particularités de la démarche évaluative. Pour être pertinente et

[8] Cette question faisait également l'objet d'un développement spécifique dans le document publié par France Stratégie en 2016, dont il constituait le premier point (« Corrélation n'est pas causalité : comment construire un contrefactuel crédible », Desplatz *et al*., 2016).

utilisable, l'évaluation doit en effet "être appropriée à son contexte et appropriée par son contexte" (Duran, 2002, p. 3-7) ». La recherche d'indépendance ne signifie pas pour autant « faire de l'évaluation sans l'administration ». Les services chargés de la mise en œuvre de la mesure évaluée disposent d'informations précieuses en la matière, et la question de leur implication dans le processus évaluatif doit donc être soulevée. L'évaluation doit également être transparente : « Conformément à la recherche de connaissance et de compréhension qui marque toute démarche évaluative, l'accès au protocole d'évaluation doit permettre de reconstituer les résultats obtenus. Le CSE [Conseil scientifique de l'évaluation] soulignait que l'évaluation doit non seulement exposer les méthodes mises en œuvre, mais "expliciter son propre mode d'emploi et ses limites" […]. Il s'agit donc aussi d'un "effort de lucidité et de réflexivité", nécessaire en raison du caractère souvent imparfait des évaluations […] »[9]. Enfin, le pluralisme de l'évaluation doit être assuré au moyen de la composition des comités de pilotage, ou de procédures garantissant l'intervention dans la démarche évaluative d'une pluralité d'acteurs, compétents sur les questions posées.

Le Conseil d'État revient plus précisément sur ces dimensions lorsqu'il évoque, de façon plus prospective, les façons de « renforcer le cycle vertueux de l'évaluation » et en particulier de « conduire une évaluation crédible, pluraliste, transparente et participative » (Conseil d'État, p. 112-117). Plusieurs propositions sont formulées dans cette perspective. L'expertise technique et administrative pourrait ainsi être plus systématiquement prise en compte pour l'évaluation des conditions de mise en œuvre de la politique évaluée. La crédibilité d'une évaluation pluraliste devrait également appeler la constitution d'une structure collective – comité de pilotage et/ou scientifique ou d'accompagnement – pour valider le cadrage de l'évaluation et les procédures de recueil et d'analyse des données. Ces comités devraient avoir une composition pluraliste, associant notamment les personnes chargées de la mise en œuvre de la politique. Des citoyens, parties prenantes et bénéficiaires de la politique évaluée pourraient également y participer. Ceci permettrait notamment aux évaluateurs de recueillir leurs perceptions des résultats intermédiaires qui leur seraient présentés, pour orienter la suite du travail évaluatif. Leur présence dans ces comités n'exonère en aucune façon le travail d'évaluation de réaliser une enquête de terrain auprès des bénéficiaires ou des partie prenantes de la politique évaluée, au moyen de méthodes qualitatives. La *démarche* évaluative comporte donc bien des exigences qui excèdent le registre de la *méthode*, et qu'il conviendrait de prendre davantage en charge.

II.2. L'ÉVALUATION BASÉE SUR LA THÉORIE

Patrice Duran préconise de « réfléchir [tant] aux conditions méthodologiques d'une bonne évaluation qu'aux dimensions contextuelles de son développement. Faire de l'évaluation un problème d'action implique […] une meilleure sensibilité au contexte d'action dans lequel elle est amenée à se déployer […]. Une vraie réflexion sur l'EPP passe nécessairement par une clarification de son usage et de ses modalités concrètes d'organisation » (Conseil d'État, 2020, p. 191). Faire de l'évaluation un problème d'action suppose notamment que celle-ci « soit une source réelle d'apprentissage [… et] il faut [pour cela] que la démarche évaluative soit intégrée aux activités de mise en œuvre » (Duran, 2018, p. 23). Appelant par exemple à une réflexion systématique sur la

[9] De ce point de vue, Adam Baïz indique dans son entretien (*Cf.* supra) que « dans la culture qualitative, les évaluateurs sont plus enclins à interroger la subjectivité des acteurs, leurs représentations, leurs 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. ».

détermination des « acteurs susceptibles d'être associés aux démarches d'évaluation » (*ibid*., p. 24), il n'explicite pas en revanche en quoi ces dernières pourraient exactement consister.

À l'instar de Patrice Duran, le Conseil d'État pointe des enjeux qu'il juge insuffisamment traités, mais n'évoque à aucun moment une démarche évaluative particulière, l'« évaluation basée sur la théorie » (TBE), qui se propose précisément de les prendre en charge. La TBE se distingue en effet dans le champ de l'évaluation en ce qu'elle constitue précisément une *démarche* évaluative, au sein de laquelle les évaluateurs conçoivent et mettent en œuvre des protocoles susceptibles d'accueillir des méthodes tant qualitatives que quantitatives. Cette démarche « s'est développée en réponse aux limites des approches expérimentales et quasi-expérimentales, qui ne permettent pas de saisir les mécanismes par lesquels une intervention produit ses impacts » (Devaux-Spatarakis, 2023a, p. 1). L'évaluation d'impact, lorsqu'elle est réalisée dans le cadre d'une TBE, explore à l'inverse le contenu de « la "boîte noire" de l'action publique en décomposant les différentes étapes de la chaîne causale liant l'intervention à ses résultats finaux » (*ibid.*). Cette explicitation de la chaîne causale constitue la « théorie d'action » de l'intervention évaluée. Elle « prend […] la forme d'une représentation graphique qui devient alors "l'échafaudage" sous-tendant les investigations de l'évaluation » (*ibid.*). C'est précisément en cela que la TBE constitue une *démarche*. Le terme de démarche, éclairé par son étymologie, renvoie à une manière de progresser, ici dans la réalisation pas à pas d'une évaluation d'impact. La démarche se distingue en cela des méthodes, ou procédés techniques qu'il conviendra de mobiliser aux différentes étapes de l'évaluation. La TBE est ainsi « susceptible de mobiliser différentes méthodes ou outils d'évaluation » (*ibid.*, p. 1). En revanche, l'explicitation de la théorie d'action constitue toujours une étape clé de la TBE. Cette théorie est aussi le truchement par lequel le travail évaluatif s'articule étroitement aux enjeux de l'action. Les évaluateurs accordent donc un soin particulier à la théorie d'action qui est « souvent co-construite entre les équipes d'évaluation et les parties prenantes » (*ibid.*, p. 5). Finement décomposée, elle présente « les principales étapes allant des réalisations aux résultats et impacts finaux en les connectant entre eux et en explicitant les hypothèses pour passer d'une étape à une autre. (…) Deux types d'investigations sont alors conduites : d'abord une analyse du changement [pour déterminer si] les étapes identifiées ou souhaitées lors de la conception de la théorie se vérifient sur le terrain […] Puis une analyse causale : comment peut-on expliquer le passage ou non d'une étape à une autre ? » (*ibid.*, p. 5). Les hypothèses de la théorie d'action de l'intervention sont testées une à une, de façon à expliquer l'atteinte ou non des effets attendus. Ces tests peuvent reposer sur des données et mobiliser des méthodes qualitatives et/ou quantitatives, en fonction des séquences analysées.

Cette démarche évaluative permet donc d'entrer dans la complexité de l'intervention évaluée, une caractéristique fréquente des interventions dans le champ des politiques de l'emploi. Elle consiste en la construction d'un plan d'évaluation ou de démonstration qui structure la conduite et l'ordonnancement des travaux depuis la collecte des données jusqu'à l'analyse. L'organisation de l'évaluation fait donc l'objet d'un investissement spécifique, dont des effets sont également escomptés en termes d'appropriation des résultats par les acteurs concernés : « Organiser l'évaluation autour de la théorie d'action est un moyen pour l'évaluateur de se concentrer autour d'une logique faisant sens pour les acteurs concernés et favorise l'appropriation des résultats ainsi que l'apprentissage qui en découlera (Weiss, 2001). » (Devaux-Spatarakis, 2014, p. 57).

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.0081.

Si l'utilisation de la TBE est longtemps restée anecdotique en France (Devaux-Spatarakis, 2014, p. 65), elle est recommandée par la Commission Européenne à partir de 2013 pour l'évaluation des fonds structurels (European Commission, 2013, p. 6)[10]. La Commission recommandait alors d'entamer toute évaluation d'impact par une analyse de la « théorie du changement » sous-tendant la politique ou la mesure évaluée, laissant ensuite ouvert le choix de la méthode retenue.

II.3. L'HÉRITAGE PRAGMATISTE

Les principes fondateurs de la TBE, formalisée dans les années 1970 et 1980 aux États-Unis, ne sont pas sans lien avec la tradition philosophique du pragmatisme. Carol Weiss, l'une des fondatrices de la TBE, partage avec le pragmatisme l'importance accordée au processus et à la conduite d'une évaluation, à ses conséquences et aux usages qu'elle rend possible, par exemple en termes d'information des processus de décision dans le champ de l'action publique (Mertens, 2015, p.79). Certains auteurs invitent à redécouvrir le pragmatisme de Peirce, Dewey et James (Dunn, 2018). Patrice Duran souligne par exemple que « si l'évaluation doit être tournée vers l'action, […] elle ne peut être une simple activité de contrôle, car, comme le déclarait le philosophe pragmatiste Charles Sanders Peirce, […] les seules conduites qui méritent vraiment d'être sous contrôle sont les conduites futures [… l'évaluation] est pleinement porteuse des questions que pose aujourd'hui le pilotage de l'action publique. C'est son intérêt, c'est sa difficulté » (Duran, 2010, p. 22).

Parfois diffus, l'héritage pragmatiste présent dans le champ de l'EPP s'explicite davantage lorsque celle-ci est mobilisée pour évaluer une expérimentation. Rien de surprenant à cela compte tenu de la place éminente de l'expérience et de l'expérimentation dans la tradition pragmatiste. Les praticiens de la TBE considèrent ainsi que parmi les différents types d'apprentissages qui peuvent résulter de l'évaluation d'une expérimentation, figure « l'apprentissage pragmatique » (Quadrant Conseil, p. 7). Celui-ci « s'inscrit dans l'idée pragmatiste que pratique et connaissance sont consubstantiellement liées : "Ce n'est que lorsqu'une solution proposée est mise en œuvre que la valeur de la solution proposée peut être établie. […] Pour acquérir le savoir, nous avons besoin d'action" (Dewey, 2005 ; Elliot, 2007 ; Gillberg *et al.*, 2011 » (*ibid.*). C'est en effet toute la question du lien entre processus d'évaluation, expérimentation, production de connaissance, identification des problèmes et action en situation que la tradition pragmatiste permet de penser et par là même d'équiper. À ce titre, la référence à la tradition pragmatique et aux travaux de Dewey pourrait trouver de nouveaux prolongements en matière d'évaluation d'impact (*infra*).

III. LA CONTRIBUTION DES MÉTHODES QUANTITATIVES ET QUALITATIVES À L'ÉVALUATION D'IMPACT

Intitulé « Comment articuler les méthodes qualitatives et quantitatives pour évaluer l'impact des politiques publiques », le guide publié par France Stratégie en 2022 est le fruit d'une collaboration avec Anne Revillard, directrice du Laboratoire interdisciplinaire d'évaluation des politiques publiques (Liepp). Prolongeant la réflexion engagée en 2016,

[10] Certains, au sein de la Commission, estimaient alors que le potentiel des « évaluations basées sur la théorie » était insuffisamment exploité (Riché, 2012)

cette nouvelle publication « vise à promouvoir une articulation à la fois rigoureuse et fructueuse des méthodes [qualitatives et quantitatives], tant auprès des institutions publiques qu'auprès des chercheurs et des praticiens du secteur privé » (De Margerie *in* Baïz *et al*., 2022, p. 4). Les auteurs y abordent de nombreuses questions essentielles relatives aux modes de production des évaluations selon les deux types de méthodes, dans le souci des singularités et apports de chacune. Ils prennent également le soin de discuter les vifs débats méthodologiques entre tenants des méthodes quantitatives acquis au raisonnement contrefactuel – qui se sont rapidement vu attribuer la plus grande rigueur scientifiques (p. 7) – et évaluateurs dits qualitativistes. Rappelant que les méthodes quantitatives ne sauraient répondre à toutes les questions évaluatives, leur contribution vient éclairer l'apport spécifique des méthodes qualitatives aux évaluations d'impact des politiques publiques. Ce guide manifeste ainsi une inflexion dans un climat jusqu'alors dominé par la référence à l'argument contrefactuel et un prima implicitement conféré aux méthodes quantitatives en matière d'évaluation d'impact.

Anne Revillard et Adam Baïz reviennent, dans leur analyse des différences entre méthodes quantitatives et qualitatives, sur l'idée selon laquelle les méthodes quantitatives visent à mesurer, quantifier et estimer l'amplitude des impacts là où les méthodes qualitatives cherchent à les expliquer. Cette ligne de partage pourrait suggérer d'associer aux uns la quantité et la mesure, aux autres la qualité et l'explication. Les auteurs prennent cependant le soin d'indiquer au préalable qu'il « ne s'agit pas d'établir une séparation nette entre ces deux types de méthodes, […] mais plutôt de caractériser ce qui distingue ces deux approches polaires […] et les spécificités idéal-typiques qu'il est raisonnable de leur associer », tout en reconnaissant qu'elles peuvent présenter des « traits communs » (Baïz *et al*., 2022, p. 13). Ceux-ci sont néanmoins peu développés dans le texte.

III.1. QUALITÉ DES QUANTITÉS ET POIDS DES QUALITÉS

La tentation d'associer, un peu rapidement, la quantité aux travaux quantitatifs et la qualité aux travaux qualitatifs semble minorer des aspects importants de chacune de ces approches.

Un enjeu essentiel de la statistique est d'attribuer une signification correcte aux objets mesurés, d'établir la « qualité des quantités » (Desrosières, 2003) et de contrôler cette qualification au fil des exploitations. Les « pertes de contrôle » sont courantes. Dans le champ des politiques d'emploi, on entend parfois désigner d'accompagnement des mesures comportant des formes de menace (de placement en formation, de suspension d'indemnité chômage, etc.) ; une corrélation positive entre accès à l'emploi et « accompagnement » n'a pourtant pas le même sens si celui-ci relève en fait de la coercition. Symétriquement, quantifier des qualités n'est pas une opération dépourvue d'enjeu analytique. La quantification ne se réduit en effet pas à la mesure. « Quantifier, c'est convenir puis mesurer » (Desrosières, 2008, p. 10), c'est « transformer des qualités en quantités […] mettre en nombre un phénomène en adoptant un étalon et en construisant un instrument », ce qui suppose « tout un travail de conception alors que la mesure s'appuie sur ce travail sans nécessairement le prendre en compte » (Moulin *et al*., 2017, p. 186-188).

À l'inverse, la tradition pragmatiste enseigne que les aspects de poids, de mesure et de pesée ne sont pas absents des travaux qualitatifs. Dans l'enquête évaluative, l'attention se focalise sur certains segments du réel, certaines réalisations et modalités d'intervention susceptibles d'être associées à des causalités. Il s'agit alors de *pondérer* l'attention que l'on accorde à un ensemble de situations qui constituent toutes a priori un « tout qualitatif

diffus » (Dewey, 1993, p. 129). Dans l'enquête qualitative en effet, « l'objet situationnel total […] est premier ; suit alors la dissociation ou la discrimination des objets mis en relation (…) Les objets ne préexistent pas à la mise en relation. Ils en émergent [… et c'est à l'enquêteur] de produire des artefacts qui [puissent rendre compte de ces] configurations de qualités [premières, d'abord perçues] » (Innis, 1998, p. 95). En outre, toute « situation indéterminée [et souvent foisonnante qui se présente dans le cours de l'enquête] est à la fois déficiente et surabondante » (Dewey, 1993, p. 279)[11]. L'enquête qualitative a pour tâche de « déterminer justement quel objet a besoin d'être éliminé, lequel a besoin d'être fourni et comment » (Dewey, 1993, 280), pour parvenir à qualifier et ainsi à clarifier chaque situation examinée. Pour Dewey, ce travail de reconstruction des situations rencontrées dans l'enquête qualitative suppose de s'assurer du « poids » de cet « objet situationnel » et des différents objets qui le peuplent comme autant « d'empiètements et d'insuffisance » (*ibid.*). Ici « la mesure […] assume donc d'abord […] une forme qualitative » (*ibid.*). Qualifier une situation foisonnante, en lien avec les questions et le problème à traiter dans l'évaluation, suppose une « mesure qualitative » (*ibid.*, p. 281), une pesée par laquelle l'enquêteur va chercher à sélectionner ou « à éliminer quelques-uns [des] facteurs [de la situation] et à introduire d'autres facteurs nouveaux » (*ibid.*, p. 296). En évaluation, ces nouveaux facteurs sont repérés ou émergent grâce aux nouvelles attentions que la théorie d'action de l'intervention évaluée ou toute autre théorie mobilisée invitent à développer. Cette « mesure qualitative », qui permet de rendre visibles et de donner un poids aux différentes composantes des situations caractérisées par l'enquête, est directement rapportée à la nature du problème à résoudre dans le cadre de l'évaluation (*ibid.*, p. 281-282). Le poids conféré, avec circonspection, aux différentes composantes de la situation dessine une configuration de dépendance mutuelle entre relations et *relata* dans le champ de pratiques observé. De là émerge un « schéma actanciel[12]» qui permet de représenter une chaîne causale intelligible. On pourrait enfin souligner que les configurations restituées à partir de l'analyse des situations associent à la fois des faits et des normes. La forme du « cas » – ou de l'étude de cas – est précisément celle où peut s'effectuer une pesée des normes *entre elles* qui, présentes de façon active dans une situation, ne sont pas nécessairement congruentes. Le cas contient alors une question, face à cette tension, celle de la norme qui peut permettre cette autre pesée et cette évaluation (Jolles, 1972, p. 142-151).

III.2. LA COMPLÉMENTARITÉ QUALI-QUANTI FACE À LA CAUSALITÉ

Adam Baïz et Anne Revilllard présentent deux conceptions de la causalité : la conception contrefactuelle des méthodes quantitatives, la conception processuelle des méthodes qualitatives (Baïz *et al.*, 2022, p. 11-18). Leurs limites respectives justifient de leur associer certaines précautions et exigences méthodologiques.

Les méthodes quantitatives permettent d'observer un impact en moyenne (*ibid.*, p. 17, 22). Les techniques des expérimentations aléatoires effectuent en effet « des tests de comparaison entre les moyennes des "effets" de l'action envisagée, sur l'échantillon traité comme sur l'échantillon témoin » (Desrosières, 2014). Elles ont ainsi « le mérite de correspondre aux besoins de l'ingénieur et de l'homme d'action, qui ont besoin de savoir "ce qui marche" avant de connaître les causes profondes des corrélations et des

[11] Dewey suggère que « la surabondance et la déficience sont aussi des concepts quantitatifs sous une forme quasi-qualitative » (Dewey, 1993, p. 280).

[12] En sociologie, le schéma actanciel mêle et traite à parité, sans prééminence de l'un sur l'autre, ses constituants de base que sont les actants et les relations, ouvrant la perspective d'une ontogénèse relationnelle.

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.0081.

réplications observées. […] cela ne devrait pas exclure de chercher à identifier des causalités plus profondes […] » (*ibid.*).

Rappelant les sources du raisonnement contrefactuel, certains auteurs soulignent que selon cette perspective il n'y a « pas de causalité sans intervention » (Smith, 2013, p. 19), l'enjeu étant bien d'identifier si l'introduction d'un moyen causatif (l'intervention) crée une différence dans les faits. Cette conception de la causalité se heurte à l'obstacle de la *préemption* (Schaffer, 2006, p. 42)[13], un obstacle fréquemment rencontré dans le champ des politiques d'emploi[14]. D'autres auteurs relèvent une distinction particulièrement utile pour situer les contributions respectives des travaux quantitatifs et qualitatifs aux évaluations d'impact : « Pour Dawid[15], il convient de distinguer l'inférence sur les effets des causes, qui consiste à comparer les conséquences attendues de différentes interventions possibles dans un système, et l'inférence sur les causes des effets, où l'on cherche à comprendre la relation causale entre un résultat déjà observé et une intervention antérieure » (Lecoutre, 2004, p. 224). La distinction entre ces deux mouvements de l'analyse inférentielle permet de marquer a priori une différence dans l'orientation respective des travaux qualitatifs et quantitatifs. Concernant « les conséquences attendues de différentes interventions possibles dans un système », les effets des causes, un raisonnement contrefactuel permet d'examiner les effets moyens de ces différentes interventions possibles (ces différentes causes possibles), maintenues chacune dans une « boîte noire ». Cela paraît problématique concernant la seconde inférence, « la relation causale entre un résultat déjà observé et une intervention antérieure », sur les causes des effets : comment remonter des effets observés à des causes, si ce n'est au moyen d'une approche processuelle retraçant a posteriori le déroulement et les composantes d'une intervention à partir d'un résultat constaté ?

Dans cette approche, « les causes et les effets sont liés par un processus continu […], la cause est la source et l'effet [en] est le réceptacle. […] la cause injecte une propriété dans l'effet » (Schaffer, 2006, p. 42). Une version détaillée de la causalité processuelle est avancée par Kistler en considérant que « la relation de responsabilité causale entre […] deux faits » (Kistler, 2006 a, p. 71) implique à la fois : (i) la présence d'un transfert ou d'une transmission de caractères entre les deux évènements – aspect mécaniste ; (ii) une relation de *dépendance nomique* entre certaines propriétés de ces deux

[13] « Un effet a plusieurs causes potentielles […et] les effets des premières empêchent les autres d'aboutir » (Schaffer, 2006, p. 42).

[14] Les travaux de la Dares sur les taux de recours et d'efficacité des différents canaux de recrutement à partir de l'exploitation de l'enquête Ofer (Offre d'emploi et recrutement) suggèrent un cas de préemption potentiellement fréquent. Ces travaux montrent en effet que « ce sont les relations personnelles ou professionnelles qui permettent le plus souvent au recrutement d'aboutir (dans 27 % des cas) devant les candidatures spontanées (21 %) » (Bergeat *et al.*, 2017, p. 4) – ainsi par exemple du « rappel de personnes ayant déjà travaillé dans l'établissement » (*ibid.*). Camille Signoretto, Aurélie Peyrin et Carine Ollivier pointent également l'importance du phénomène de réembauche (Signoretto *et al.*, 2023-2024), et signalent que l'Unédic parle de « relation de travail suivie » (Journeau, 2019, p. 2) pour caractériser la relation entre un salarié et un employeur conduisant à une série de réembauches, sous certaines conditions (Signoretto *et al.*, 2023-2024, p. 16). A la lumière de ces résultats, un cas de préemption potentiel est le suivant : le chemin causal établi par définition, c'est-à-dire conçu selon une « causalité institutionnelle » (Gislain, 2002), qui conduit un bénéficiaire d'une modalité d'accompagnement du service public de l'emploi jusqu'à l'accès à l'emploi, peut ne pas être effectué jusqu'à son terme – ne pas aboutir – parce qu'il est interrompu par cet évènement qu'est le rappel et la réembauche par un ancien employeur de ce bénéficiaire en cours d'accompagnement. Cet évènement (cette autre cause) amorce un autre chemin causal, concurrent du premier, celui d'une « relation de travail suivie » selon la qualification de l'Unédic qui, initié par les besoins contingents d'un ancien employeur, permet au bénéficiaire d'accéder à l'emploi (réembauche).

[15] A propos de (Dawid, 2000).

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.0081.

évènements – aspect explicatif (*ibid*., p. 72) [16]. Kistler introduit la notion de responsabilité causale dans des développements consacrés aux sciences de la nature ; il ne prétend pas que sa solution puisse être étendue aux sciences sociales. Pour ce faire, il faudrait que le principe de « responsabilité causale » prenne un tour plus contextuel et/ou moins universel que dans la tentative d'unification conceptuelle proposée par Kistler (Kistler, 2006b). Au plan méthodologique, on pourrait parfaitement conserver la double préoccupation qui était la sienne : repérer la transmission d'un caractère entre évènements ; mobiliser, pour l'interprétation, une dépendance habituellement renseignée entre certaines propriétés de ces événements. Il ne serait plus question, pour autant, de concevoir la causalité comme relation réelle, objectivement, mais de donner un tour plus « conventionnel » et/ou pragmatiste à l'application du principe du causalité : la recherche d'une responsabilité causale ne serait plus qu'une façon de poser, par référence à la théorie d'intervention, des questions auxquelles il appartiendrait au « réel » de répondre ou pas.

IV. LES ÉTUDES DE CAS

L'étude de cas constitue une sorte d'allant de soi des travaux qualitatifs. L'usage de ce terme n'est pourtant pas exempt d'un certain nombre d'ambiguïtés.

Méthode transversale, elle consiste, selon Anne Revillard et Adam Baïz, à réaliser une étude approfondie d'un sujet précis et limité pour représenter les nuances et la complexité d'un comportement, d'un phénomène ou d'une situation. L'étude de cas a beau par définition porter sur une situation unique, elle « peut fournir un faisceau d'indices sur l'existence d'un impact, son amplitude et ses ressorts. […] Au regard de la nature des événements et de leur enchaînement, [elle] pourra raisonnablement inférer un lien de cause à effet » entre deux évènements (Baïz *et al*., 2022, p. 30). Ainsi d'une nouvelle réglementation et de la chute de la pollution pour reprendre un exemple des auteurs. Les données collectées dans l'étude de cas aident à formuler une hypothèse (indices), à conforter une théorie (preuves) ou au contraire à la réfuter (contre-exemples) et sert donc directement l'analyse causale (*ibid*., p. 30).

IV.1. LE CAS N'EST PAS UNE ILLUSTRATION

Monographie et étude de cas sont parfois confondues. Le flou qui entoure le premier de ces termes a pu contribuer au flou caractérisant en retour le statut des cas dans l'approche monographique.

Anne Revillard et Adam Baïz lèvent d'emblée l'ambiguïté : la monographie est le nom que prend la restitution des données collectées et leur mise en perspective (*ibid*., p. 30). Ils écartent ainsi une tendance chez les commanditaires publics à parler de monographies pour se référer aux différents lieux de l'étude évaluative (monographies territoriales, d'entreprises, etc.). Cette tendance est particulièrement dommageable

[16] Kistler propose, dans un premier temps, que les causes sont impliquées de manière processuelle dans les effets en ce qu'elles leur « injectent » une propriété : la causalité s'analyse alors comme transfert de propriété à travers une chaîne d'événements précisément capable de transmettre un « caractère ». Pour établir une « responsabilité causale » – notion qu'il préfère à celle de causalité – Kistler « double » cette première conception de la causalité comme transfert par la prise en considération, par ailleurs, d'une dépendance connue – qu'il appelle « dépendance nomique » (au sens d'une loi ou d'une norme établies) – entre certaines des propriétés des événements à expliquer : la relation de responsabilité causale s'établit alors entre des faits qui trouvent précisément à instancier l'une ou l'autre de ces propriétés.

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.0081.

lorsque la monographie est par ailleurs la seule forme de restitution attendue dans le cahier des charges d'une évaluation. Le risque est alors que les études de cas proprement dites se dissolvent dans des monographies locales, dans lesquelles la mobilisation des cas ne relève plus que du registre de l'illustration ou de l'exemple. Les cas y sont alors cités de façon ponctuelle et un peu dispersée. Simultanément, le risque est que les études de cas cèdent la place à la restitution de pratiques « moyennes », d'entreprise par exemple, qui constitueront le cœur de la monographie. Il manque alors ce qui fait l'essentiel de l'étude de cas et de son apport pour l'évaluation : la possibilité d'inférer des liens de cause à effet tout au long de la reconstruction d'un processus réel, peuplé de conditions particulières à identifier, ponctué d'évènements dont il faut restituer l'enchaînement pour parvenir à reconstituer un chemin causal, les effets intermédiaires et l'impact final dans le cas étudié, ainsi que l'indiquent Adam Baïz et Anne Revillard.

En lien avec le flou qui entoure la notion de monographie, les critères de sélection des lieux retenus pour réaliser les investigations de terrain (et produire les monographies « locales ») invoquent généralement la nécessité de disposer d'une diversité de contextes territoriaux ou d'entreprises, dûment caractérisés. Hors l'étape de sélection des terrains, il n'est cependant pas toujours « fait cas » dans l'analyse proprement dite de tels contextes, rapidement renvoyés au statut de décor. Il importe, à l'inverse, de sélectionner les cas en fonction du problème à résoudre de sorte à saturer la diversité des pratiques observées dans l'enquête. C'est à cette diversité que les études de cas doivent faire droit, en analysant les différentes configurations ou combinaisons de facteurs ayant permis ou non de produire les effets escomptés. Certains éléments du contexte peuvent alors être intégrés à l'analyse, au titre de conditions dont la présence contribue à la production des effets.

Le commentaire des études de cas présenté par Adam Baïz et Anne Revillard aurait pu insister davantage sur l'importance de la finesse du grain et du niveau de détail dans lequel il convient d'entrer pour accéder, dans chaque cas examiné, à « ce qui se transmet » et dont il faut repérer les indices et chercher la trace (Ginzburg, 2010). C'est tout l'enjeu d'une analyse suffisamment fine des situations traversées par les personnes (ou les entreprises), d'une analyse des fonctionnements que ces situations favorisent, des apprentissages et des développements qu'elles soutiennent et font éclore ou qu'au contraire elles entravent. L'étude de cas est le lieu privilégié de cette forme d'analyse. Il y aurait un intérêt fort à affirmer plus nettement les spécificités du raisonnement à partir de cas singuliers et le type de données qu'il suppose d'obtenir. Cela permettrait, dès lors que l'étude de cas est invoquée dans le cadre des travaux qualitatifs, de marquer davantage l'écart avec le traitement des cas dans les travaux quantitatifs.

Ainsi, dans l'un des passages consacrés à la comparaison entre méthodes quantitatives et qualitatives du point de vue de leur usage respectif des cas, les auteurs évoquent l'exemple d'une évaluation de l'impact d'une réforme sur une population de deux millions d'entreprises (Baïz *et al*., 2022, p. 22). De leur point de vue, une méthode quantitative cherchera à observer un échantillon représentatif de la population d'ensemble, tandis qu'une méthode qualitative « caractérisera précisément une poignée d'entreprises (le chiffre d'affaires, la taille, la nature des relations interpersonnelles, etc.) » (*ibid.*). Elle pourra y inclure des caractéristiques inobservables pour le statisticien, telles que la nature des relations interpersonnelles entre les salariés par exemple. Cette brève évocation – qui ne cherche certes pas à être exhaustive – des caractéristiques recherchées pour chaque entreprise de la « poignée » sélectionnée reste cependant proche des types de caractéristiques susceptibles d'être identifiés dans le cadre de travaux

quantitatifs. Dans une perspective évaluative, on aurait par exemple pu s'attendre à ce que soient évoquées des données qualitatives sur l'intention ou la stratégie développée par les entreprises investiguées, en lien avec la mise en œuvre de l'intervention, au positionnement et aux fonctionnements qu'elles ont choisis d'adopter. Selon les cas, ces fonctionnements peuvent par exemple être conformes à l'esprit de l'intervention, ou contribuer à son détournement, son évitement, etc. Ces différentes dimensions, auxquelles le travail qualitatif peut précisément donner accès, sont essentielles. En leur absence, on peut avoir le sentiment que le parallélisme recherché dans l'exposition des méthodes quali- et quantitatives, à des fins de comparaison, les maintient dans une sorte de proximité plutôt qu'il ne témoigne de leurs singularités respectives et de l'écart qui les sépare dans leur rapport aux « cas » et à leur exploitation. De plus, ces autres dimensions, relatives à l'intention ou la stratégie peuvent marquer singulièrement les points d'impact ou d'accroche par lesquels une intervention peut intéresser et rencontrer son public, et l'usage que celui-ci pourra en faire par la suite.

On pourrait ainsi considérer que la spécificité des travaux qualitatifs tient à la façon dont ils s'intéressent tant aux caractéristiques des personnes ou des entreprises destinataires d'une mesure, qu'à la *situation* dans laquelle elles sont d'abord inscrites et à la *façon dont celle-ci évolue progressivement* au bénéfice de l'intervention. Les acteurs forment des intentions, adoptent certaines orientations normatives (ou « normes d'action »), *en fonction de leur situation* lorsqu'intervient la mesure. Ils font d'un certain point de vue une expérience toujours singulière du dispositif ou du cadre expérimental qui leur est adressé. Et cette expérience est toujours cadrée par la situation dans laquelle elle s'amorce et peut se déployer. L'attention à la situation, une dimension souvent « négligée » (Quéré, 1997)[17], semble particulièrement féconde pour comprendre comment et pourquoi une mesure de politique publique rencontre ou non son public (entreprises, personnes bénéficiaires, etc.) *en situation*, et parvient ou non à produire pas à pas des effets. L'attention par exemple à l'évolution des situations traversées par des bénéficiaires d'un dispositif public d'accompagnement peut singulièrement renforcer l'identification des « causalités intermédiaires », qui constituent comme Anne Revillard et Adam Baïz le soulignent (p.15) les maillons essentiels à l'établissement d'une « causalité processuelle ».

Depuis 2022, Anne Revillard a prolongé ces réflexions en coordonnant un ouvrage collectif dédié aux « méthodes et approches en évaluation des politiques publiques » (Revillard, 2023). L'ouvrage présente un certain nombre de méthodes et d'approches qualitatives, quantitatives et mixtes, selon une trame commune facilitant leur comparaison. Il donne à voir, de façon plus précise que ne le fait le document France Stratégie, le large éventail des approches disponibles en matière d'EPP. L'étude de cas, conçue ici comme stratégie de recherche plutôt que comme méthode, fait l'objet d'un chapitre spécifique (Ridde *et al.*, 2023). Celui-ci rappelle que les études de cas peuvent mobiliser des méthodes qualitatives, quantitatives ou mixtes. Les études de cas sont aussi très présentes dans d'autres chapitres de l'ouvrage consacrés à plusieurs méthodes qualitatives ou mixtes : ainsi des « cas réels ou typiques » mobilisés dans le traçage de processus (Raimondo, 2023b), des « configurations issues d'études de cas » de

[17] Cette dimension n'est pas absente du document d'Anne Revillard et Adam Baïz, qui soulignent que « les méthodes qualitatives se proposent d'aborder les individus, les situations et les phénomènes, non pas en les réduisant sur des indicateurs chiffrés, mais en mobilisant des mots (concepts) et des structures de mots formant des descriptions et des raisonnements » (p. 20). Mais l'analyse spécifique et précise des situations, de leur « aménagement », leur « potentiel » et leur évolution, ne fait pas l'objet de développements particuliers dans les parties consacrées aux méthodes qualitatives.

l'évaluation réaliste (Louart *et al*., 2023), ou des « tests empiriques réalisés dans des études de cas » et des « récits de contribution » dans l'analyse de contribution (Delahais, 2023). Il est précisé que ces trois dernières approches relèvent toutes d'une démarche d'évaluation basée sur la théorie, qui fait également l'objet d'un chapitre dédié (Devaux-Spatarakis, 2023b). Les chapitres consacrés au traçage de processus et à l'évaluation réaliste, en particulier, permettent d'appréhender plus précisément la façon dont ces approches s'appuient sur des études de cas pour déterminer les conditions dans lesquelles certains types d'effets peuvent être produits (traçage de processus), ou pour identifier les mécanismes qui, déclenchés par la politique évaluée dans des contextes spécifiques, vont produire des effets (évaluation réaliste). Le chapitre consacré aux études de cas confère aux cas différents statuts possibles justifiant leur usage dans une évaluation : « cas critiques, uniques, typiques, révélateurs, instrumentaux » (Ridde *et al.*, 2023, p. 2). Il détaille ensuite la place respective, dans l'étude de cas, des « données empiriques et [de] la théorie, qu'elle soit la théorie d'intervention de la politique ou un cadre théorique ou conceptuel utilisé pour comprendre la politique. Ce processus est nommé par Yin la généralisation analytique » (*ibid.*, p. 2).

IV.2. QU'EST-CE QU'UN CAS ?

S'ils contribuent à clarifier la discussion sur l'étude de cas, Anne Revillard et Adam Baïz laissent de côté des questions de fond : « qu'est-ce qu'un cas ? » (Ginzburg, 2007, p. 44), « de quoi mon cas est-il le cas ? » (Ragin *et al*., 1992), comment l'établir « comme cas de quelque chose » ? (Dodier, 1995, p. 125). Ces questions sont essentielles pour décrire plus précisément le double mouvement que suppose l'étude de cas : le regard « au microscope » (Ginzburg, 2013) qui « piste » et produit les données du cas, et le travail de condensation conceptuelle, dont découlera sa portée. Ces deux mouvements sont distincts, mais ils s'articulent et se répondent tout au long de l'étude de cas. L'héritage pragmatiste est particulièrement fécond sur ce point. L'enquête, pour Dewey, est la transformation dirigée d'une situation indéterminée en une situation unifiée ; elle est la détermination progressive d'un problème (Dewey, 1993). Dans l'enquête scientifique et en évaluation, des hypothèses, des idées, des concepts y sont nécessaires, au titre à la fois de « moyens et [de] fins possibles de solution » (*ibid*., p. 183) tout au long de l'examen des situations et des cas particuliers soumis à enquête. La mobilisation des idées et hypothèses comme *moyens* et comme *fins* renvoie à deux mouvements distincts, quoiqu'étroitement liés, du travail à réaliser dans une étude de cas.

Le regard « au microscope » identifie d'abord, dans les données obtenues par l'enquête qualitative, les traces et les indices de l'intervention évaluée et de ses effets. Il faut par exemple chercher la trace, dans les entretiens avec un bénéficiaire d'un dispositif d'accompagnement et de son conseiller, des modalités précises d'accompagnement effectivement mises en œuvre, de leur évolution tout au long du parcours, des occasions qui ont été – ou non – proposées à ce bénéficiaire (périodes de mise en situation professionnelle, formation, intégration dans une communauté de travail, etc.), des expériences qu'elles lui ont permis de faire, de la façon dont le conseiller les a – ou non – accompagnées, des capacités que la personne a pu développer, etc. Ces traces doivent se lire dans l'étendue foisonnante d'une série de situations qualitatives, encore mal déterminées. Le cas n'est pas donné d'emblée. Il n'est pas une illustration ou un exemple. Il ne se présente pas comme un ensemble d'éléments immédiatement visibles, observables et détachables les uns des autres. Certaines situations, certaines expériences sont d'abord confuses : elles posent problème sans que le problème soit encore formulable ; elles trouvent une issue favorable sans qu'il soit possible de comprendre

pourquoi. Elles supposent d'être éclaircies en formant la matière d'un cas. Le cas, dans cette perspective, se compose et se produit, en analysant finement tout ce dont il est fait.

Des moyens y sont nécessaires. En évaluation, produire le cas, repérer les traces et les données pertinentes suppose de « comparer, de manière systématique et rigoureuse, les données empiriques et la théorie » (Ridde *et al.*, 2023). Les hypothèses de la théorie d'action de l'intervention évaluée et de toute autre théorie mobilisée sont utilisées comme *moyens* permettant de prêter attention à des faits qui auraient pu passer inaperçus, à des conditions qui auraient pu être considérées comme évidentes, de peu d'importance, etc. Les idées et les hypothèses sont toujours d'abord « opératives » (Dewey, 1993) : elles provoquent et dirigent les observations qui « fournissent une nouvelle matière factuelle » (*ibid.*, p. 183) tout au long du travail de composition du cas. Elles sont une première médiation nécessaire à l'identification de données qui ne seraient pas immédiatement accessibles, à la qualification des faits jugés pertinents dans un enchaînement causal pas à pas recomposé. Ce travail s'ouvre également à des données inattendues au regard des hypothèses et des objectifs du programme évalué. Ces autres données peuvent apparaître au moyen de la mobilisation d'autres hypothèses, qui n'avaient pas – encore – leur place dans la théorie d'action de la mesure évaluée.

À mesure que le cas se compose, les situations indéterminées qui se présentent sont circonscrites, clarifiées, qualifiées. La nécessité de produire les données pertinentes du cas s'articule progressivement à un autre impératif : comprendre peu à peu « de quoi ce cas s'avère être le cas ». Dans cet autre mouvement, les idées, hypothèses et concepts vont occuper une autre fonction. La composition du cas s'épure, à mesure que les données obtenues sont soupesées et que sont sélectionnées celles qui doivent être conservées. Elle se densifie également, à mesure que des idées, un ou des concepts parviennent à le qualifier ou à qualifier ses principales composantes, à le caractériser en une juste problématisation, une juste formulation du problème dont il s'avère *in fine* être le cas. Le cas et « la chose dont il est le cas » s'élaborent ainsi selon deux mouvements, étroitement imbriqués : les études de cas doivent identifier et présenter les données utiles et nécessaires à la structure du problème à *construire* progressivement, et ainsi à « résoudre ». C'est tout l'enjeu du travail de conceptualisation du cas. Dans ce travail, les idées, les hypothèses et concepts mobilisés représentent désormais des « significations hypothétiques possibles » (*ibid.*, p. 628), des modes et des « *fins* possibles de solution » (*ibid.*, p. 183) de la situation qui, d'indéterminée, se problématise progressivement. Ces significations et solutions sont mises, une à une, à l'épreuve des données du cas. Le « terrain », les données du cas répondent – ou non – aux significations et conceptualisations proposées. Dewey considère que les concepts et idées qui fonctionnent comme résolution possible d'un cas et qui, finalement, le qualifient, ont une « capacité représentative » (*ibid.*, p. 628). En effet ces idées « peuvent être mises pour une solution possible » (*ibid.*) de la situation indéterminée soumise à enquête dans le cadre de l'étude de cas. Elles la problématisent. Le cas « exige […in fine] d'être condensé, réduit à l'essentiel, c'est-à-dire porté jusqu'à son concept. Le concept est relatif au cas ; il permet […] de déterminer le cas de quoi serait le cas pourrait-on dire en paraphrasant Charles C. Ragin, et a pour fonction de le problématiser, d'en concevoir la carte, en réalité de le résoudre à travers sa capacité à le condenser » (Duclos, 2024, p. 14). La « chose dont il est le cas » constitue le résultat, le produit conceptuel de l'étude de cas. Le cas, lorsqu'il est ainsi condensé en un ou plusieurs concepts en interrelation, est représentatif du problème qu'il a permis progressivement de déterminer, et sous cette forme de résoudre.

L'étude de cas peut faire émerger une *configuration* (Ridde *et al.*, 2023)[18]. Le cas a une dimension temporelle[19]. Il se déploie et produit en son sein des conséquences. La composition, la problématisation et la conceptualisation du cas permettent d'identifier les effets et d'établir les causalités intermédiaires qui le ponctuent, jusqu'à l'impact final, terme du processus d'intervention singulier qu'il a permis de conceptualiser.

Cette façon de concevoir le cas permet de penser différemment sa portée, son usage et sa possible transférabilité à d'autres cas, d'autres situations.

IV.3. TRANSPORT ET « TRANS-FAIRE »

La transférabilité, opérant de cas singuliers en cas singuliers, prend la forme d'un transport de cas en cas, distinct en cela des modalités de la généralisation et de l'extrapolation dans les travaux quantitatifs.

Anne Revillard et Adam Baïz font de l'extrapolation un procédé commun aux méthodes quanti- et qualitative, avec des variations. De leur point de vue, le quantitatif permet une extrapolation moyenne, et le qualitatif une extrapolation ciblée des résultats qu'il produit. On peut cependant se demander si l'effort de symétrie, induit par la comparaison systématisée des deux méthodes, ne peut pas conduire à atténuer là aussi certaines spécificités concernant spécialement la façon dont le travail qualitatif sur des cas permet d'envisager la question de la transférabilité/transposabilité des résultats évaluatifs obtenus.

Les auteurs illustrent le procédé de l'extrapolation ciblée, attribué aux méthodes qualitatives, par un exemple d'évaluation évoqué dans la première partie du document – sans que l'ensemble de cette évaluation soit présenté (Baïz *et al*., 2022, p. 22) : dans une enquête qualitative, les résultats observés sur les quelques entreprises interrogées ne sont pas a priori extrapolables à l'ensemble de la population, du fait de la petitesse de l'échantillon observé. Ils pourraient en revanche, selon eux, être raisonnablement extrapolés à toute entreprise présentant le plus précisément possible les mêmes caractéristiques que l'une des entreprises observées. En d'autres termes, si deux entreprises partagent un grand nombre de caractéristiques, ce qu'une méthode qualitative vise précisément à observer, il est très probable que l'impact observé chez l'une soit extrapolable chez l'autre. Du point de vue des auteurs, cette extrapolation ciblée, et non en moyenne comme le ferait une analyse quantitative, permet d'appréhender des cas particuliers. Dans cette perspective, tout se passe un peu comme si le qualitatif continuait d'emprunter ses modes de raisonnement au quantitatif, mais « en plus petit » (sur une plus petite échelle, « une poignée d'entreprises », un « nombre limité de cas »), et en mobilisant pour chaque cas un plus grand nombre de caractéristiques (plus détaillées, observables et inobservables). Le terme d'extrapolation pose problème concernant des méthodes qualitatives, en les maintenant encore en proximité des méthodes quantitatives et de leurs procédés. L'extrapolation vise à combler des vides. On extrapole un résultat à quelque chose que l'on ne connaît pas. Le qualitatif procède autrement.

Les auteurs auraient par exemple pu s'appuyer sur les développements des sciences sociales autour de la « pensée par cas » (Passeron *et al*., 2005), pour creuser davantage la spécificité des travaux qualitatifs en la matière. Dans le travail qualitatif, la transposition, avec ses variations, s'opère de cas à cas, plutôt que d'un cas vers une totalité identifiée,

[18] Voir aussi la configuration « contexte, mécanismes, effets » en évaluation réaliste (Louart *et al.*, 2023).
[19] Cf. le bénéfice d'une « approche diachronique afin de produire des études de cas longitudinales » (Ridde *et al.*, 2023., p. 183).

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.0081.

ainsi que procède la généralisation dans les travaux quantitatifs à partir d'un échantillon représentatif. Comme le souligne Bruno Latour, « on ne passe pas du [cas] particulier au [cas] général [et donc à un tout], mais du [cas] particulier à d'autres [cas] particuliers » (Latour *et al.*, 2013, p. 211). L'étude de cas permet de déterminer et de produire les éléments susceptibles de se transporter. Elle élabore « la chose dont le cas analysé est le cas » et qui peut être « répétée avec des variations » (*ibid.*, p. 225). Une configuration par exemple, comme suite de relations entre des évènements dotés de certaines propriétés[20], peut être à la fois « suffisamment répétée pour être identifiée comme étant la même ; suffisamment variée pour être transposée plus loin dans le temps et l'espace » (*ibid.*), connectable, avec des variations, à des environnements différents. Une séquence spécifique d'un processus d'accompagnement par exemple, qu'une étude de cas aura permis de caractériser et de conceptualiser, peut être transportée en inspirant de nouvelles manières de faire, en lien avec de nouveaux environnements. Le degré de conceptualisation des différentes séquences et composantes du processus évalué leur confère la flexibilité nécessaire à leur usage dans d'autres cas. Le transport dans un autre cas peut aussi s'engager modestement, à travers un « acte initial qui diffère de l'habituel » (Dewey, 2023, p. 38). Cet acte, inspiré de réalisations concrètes valorisées dans l'évaluation, constituera « l'entame d'une série d'actes susceptibles de […] faire adopter [pas à pas une nouvelle] habitude » (*ibid.*), un nouveau fonctionnement. Dans les travaux qualitatifs le transport prend la forme d'un « trans-faire », selon la belle expression de Cynthia Fleury et Antoine Fenoglio (Fenoglio *et al*., 2022, p. 13). Les travaux qualitatifs sont « moins concernées par la question de savoir dans quelle mesure une intervention a eu un impact sur un résultat souhaité » – à l'instar des travaux quantitatifs – ; ils s'intéressent à la question de savoir « dans quelles conditions, comment et pourquoi une intervention a fonctionné dans le monde réel » (Raimondo, 2023a, p. 3). Des conditions mises à jour par l'évaluation qualitative peuvent être transportées et éprouvées dans d'autres situations, avec des variations. Il ne s'agit pas ici de combler les vides – par extrapolation – que représenteraient les autres cas à venir. Les études de cas ont permis de décrire et de suivre des processus réels, de comprendre la façon dont une intervention a composé et fonctionné avec son environnement pour produire les effets observés dans les différents cas analysés. Il y a à suivre les processus d'appropriation et de transport de certaines séquences de l'intervention évaluée, de concepts et de problèmes mobilisables dans de nouvelles situations et pour de nouveaux cas. Il y a à observer la capacité de ces « éléments transportés » à entrer en transaction avec de nouveaux environnements, une capacité dont ils ont précisément déjà fait preuve dans les cas examinés par l'évaluation. Il y a à observer la façon dont les problèmes élaborés par les études de cas résonnent avec de nouvelles situations à venir, conservent une capacité à les éclairer, les problématiser et les transformer.

Pour finir, les études de cas constituent des produits susceptibles d'être facilement appropriables, discutables par les parties prenantes de la mesure ou de la politique évaluée. Elles se présentent sous des formats qui donnent prise aux différents acteurs impliqués dans la mise en œuvre de la mesure, en qualifiant, en problématisant et en conceptualisant des situations qui, d'une façon ou d'une autre, feront écho à celles qu'ils ont à prendre en charge. Il en va de même du côté des bénéficiaires de la mesure, dont des expériences singulières sont également retracées. En cela, les études de cas peuvent être mises en partage et en débat dans le cadre d'un « espace de parole pluraliste » (Terzi, 2021, p. 85), dans le cours du travail évaluatif, qui viendrait ainsi renforcer la mise à

[20] Pour reprendre les termes de Kistler (Kistler, 2006 a).

l'épreuve de « la force probatoire des propositions » (*ibid.*, p. 81) émises par les évaluateurs, dans le prolongement de la tradition pragmatiste. Des temps de restitution intermédiaire, par exemple, pourraient être investis dans cette perspective. Le travail d'administration de la preuve et la mise en débat des propositions évaluatives s'enrichiraient ainsi des « extériorités qu'elles convoquent et auxquelles elles [ouvriraient] un espace de parole » (*ibid.*, p. 87). Il y a sans doute là matière à réfléchir aux types de produits susceptibles d'être appropriables dans le cadre d'une évaluation pluraliste, et aux types de démarches évaluatives qui pourraient effectivement la soutenir.

Enfin, penser ainsi la question de la transférabilité des propositions issues d'une évaluation supposerait également de s'interroger sur les *a priori* des différentes approches évaluatives. Ainsi de l'hypothèse de scalabilité (Tsing, 2021) sur laquelle repose *a priori* « l'expérimentation pour l'évaluation », là où « l'évaluation par l'expérimentation » (Duclos, 2024) et les travaux qualitatifs permettent de travailler sous l'hypothèse inverse, celle de la non-scalabilité des programmes, *a fortiori* dans le champ des politiques d'emploi où les effets contextuels sont importants.

RÉFÉRENCES BIBLIOGRAPHIQUES

Aguilera T., Chevalier T., 2021, « Les méthodes mixtes pour la science politique. Apports, limites et propositions de stratégies de recherche », *Revue française de science politique*, vol. 71, p. 365-389.

Baguelin O., Duclos L., 2023, « Les politiques de l'emploi au régime de l'expérimentation : la preuve de concept », *Socio-économie du travail*, n°14, 2023, p. 17-54.

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.

Baïz, A. (avec la contribution de Guyot M., Lewandowski M., Suty A.), 2022, Quelles évaluations des politiques publiques pour quelles utilisations ? Rapport, France Stratégie.

Bono P.-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C., Trannoy A., 2018, « Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique », Document de travail, France Stratégie, n°2018-06.

Card D., Kluve J., Weber A., 2015, « What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations », NBER Working Paper, n°21431.

Concato J., Shah N., Ralph I., Horwitz R., 2000, « Randomized, controlled trials, observationnal studies, and the hierarchy of research designs », *The New England Journal of Medicine*, 342(25), p. 1887-1892.

Conseil d'analyse économique, 2013, « Evaluation des politiques publiques », *Les notes du conseil d'analyse économique*, n°1.

Conseil d'État, 2019, Les expérimentations : comment innover dans la conduite des politiques publiques, Les études du Conseil d'État, La documentation Française.

Conseil d'État, 2020, Conduire et partager l'EPP, Les études du Conseil d'État, La documentation Française.

Bergeat M., Remy V., 2017, « Comment les employeurs recrutent-ils leurs salariés ? », *Dares Analyses*, n°064.

Dawid A.P., 2000, « Causal inference without contrefectuals (with comments and rejoinder) », *Journal of the American Statistical Association*, 95, 450, p. 407-448.

Delahais T., 2023, « L'analyse de contribution », in Revillard A. (éd.), 2023, Méthodes et approches en EPP, Québec, Editions science et bien commun, p. 293-305.

De Margerie G., 2022, « Avant-propos » in Baïz A., Quelles évaluations des politiques publiques pour quelles utilisations ? Rapport, France Stratégie.

Desplatz R., Ferracci M., 2016, Comment évaluer l'impact des politiques publiques ? un guide à l'usage des décideurs et des praticiens, France Stratégie.

Desplatz R., Lacouette-Fougère C., 2019, « L'EPP en France », Document de travail, France Stratégie, n°2019-13.

Desrosières A., 2003, « Les qualités des quantités », *Courrier des statistiques*, n°105-106, p. 51-63.

Desrosières A., 2014, « Deux sens de l'évaluation : les expérimentations aléatoires », in Desrosières A. (éd.), 2014, *Prouver et gouverner. Une analyse politique des statistiques publiques*, Paris, La Découverte, p. 60-69.

Devaux-Spatarakis A., 2014, « L'évaluation 'basée sur la théorie', entre rigueur scientifique et contexte politique », *Politiques et management public*, vol. 31/1, p. 51-68.

Devaux-Spatarakis A., 2023a, « L'évaluation basée sur la théorie », Fiche méthodologique, LIEPP, n°27.

Devaux- Spatarakis A., 2023b, « L'évaluation basée sur la théorie », in Revillard, A. (éd.), 2023, Méthodes et approches en EPP, Québec, Editions science et bien commun, p. 269-278.

Dewey J., [1967] 1993, *Logique : la théorie de l'enquête*, Paris, PUF.

Dewey J., 2005, *The quest for certainty: a study of the relation of knowledge and action*, New York, Minton, Balch & company.

Dewey J., [1930] 2023, *Nature humaine et conduite. Introduction à la psychologie sociale*, Paris, Gallimard.

Dodier N., 1994, « Ragin, C. C., Becker, H. S., What is a case ? Exploring the foundations of social inquiry » [Compte-rendu], *Revue française de sociologie*, n°35-1, p. 125-128.

Duclos L., 2024, « L'évaluation par l'expérimentation », Document d'études DGEFP-IDHES, ministère du Travail, décembre.

Dunn W. L., 2018, « Rediscovering Pragmatism as the Origin of the Policy Sciences », *European Policy Analysis*, vol. 4, n°1, p. 13-22.

Duran P., 2002, « L'évaluation, au risque de son institutionnalisation », *Politiques et management public*, vol. 20, n°4, p. 3-7.

Duran P., 2010, « L'EPP : une résistible obligation. Introduction générale », *Revue Française des Affaires Sociales*, n°1-2, p. 7-24.

Duran P., 2018, « L'EPP : les sciences sociales comme sciences de gouvernement », *Idées économiques et sociales*, n° 193, p. 6-27.

Duran P., Ehrel C., Gautié J., 2018, « L'évaluation des politiques publiques : les sciences sociales à l'épreuve », *Idées économiques et sociales*, n°193, p. 4-5.

Duran P., 2021, « Evaluation des politiques publiques : les leçons de l'expérience », *Revue française d'administration publique*, n°177, p. 1-15.

Elliott J., 2007, *Reflecting Where the Action Is : The selected works of John Elliott*, London and New York, Routledge Publishers.

European Commission, 2013, « Approach and Methodology », in European Regional Development Fund and Cohesion Fund, 2013, Guidance for the Terms of Reference for Impact Evaluations the Programming Period 2014-2020.

Gillberg C., Vo L.C., 2011, « Approche pragmatiste de la connaissance et de l'apprentissage dans les organisations », *Management & Avenir*, n°43, p. 410-427.

Ginzburg C., 2007, « Réflexions sur une hypothèse vingt-cinq ans après », in Thouard, D. (éd.), 2007, *L'interprétation des indices. Enquête sur le paradigme indiciaire avec Carlo Ginzburg*, Villeneuve d'Ascq, Presses universitaires du Septentrion.

Ginzburg C., 2010, *Mythes, emblèmes, traces. Morphologie et histoire*, Paris, Verdier.

Ginzburg C., 2013, « Nos mots et les leurs. Une réflexion sur le métier de l'historien, aujourd'hui » [Traduction de Martin Rueff], ESSAIS, Hors-série 1, p. 191-210.

Gislain J.-J., 2002, « Causalité institutionnelle : la futurité chez J.R. Commons », *Économie et institutions*, n°1, p. 47-66.

Innis R.E., 1998, « John Dewey et sa glose approfondie de la théorie peircienne de la qualité », *Protée*, vol. 26, n°3, p. 89-98.

Jolles A., 1972, *Formes simples*, Paris, Seuil.

Journeau F., 2019, « Relations de travail suivies : 2 CDD courts sur 3 s'inscrivent dans une succession de contrats courts chez le même employeur », Éclairages, Unédic, juillet.

Kistler M., 2006a, « La causalité comme transfert et dépendance nomique », *Philosophie*, n°89, p. 53-77.

Kistler M., 2006b, « Unité des sciences, causalité et rôle causal de l'esprit », *Linx*, n°54, p. 139-150.

Latour B., Jensen P., Venturini T., Grauwin S., Boullier D., 2013, « 'Le tout est toujours plus petit que ses parties' Une expérimentation numérique des monades de Gabriel Tarde », *Réseaux*, n°177, p. 197-232.

Lecoutre B., 2004, « Expérimentation, inférence statistique et analyse causale », *Intellectica*, n°38, p. 193-245.

Louart A, Habibata B., Robert E., Ridde V., 2023, « Evaluation réaliste », in Revillard, A. (éd.), 2023, Méthodes et approches en EPP, Québec, Editions science et bien commun, p. 279-291.

Mertens D.M., 2015, "Philosophical assumptions and programme evaluation", *SpazioFilosofico*, 13, p. 75-85.

Moulin S., Beaud J.-P., 2017, « Quantification et mesure », in Bouchard F., Doray P., Prud'homme J. (éd.), 2017, *Sciences, technologies et société de A à Z*, Presses de l'Université de Montréal, p. 186-188.

Passeron J.-C., Revel J., (éd.), 2005, *Penser par cas*, Paris, Éditions EHESS.

Quadrant Conseil, 2020, « Évaluer les expérimentations : Enjeux et postures de l'évaluateur/rice ».

Quéré L., 1997, « La situation toujours négligée ? », *Réseaux*, vol. 15, n°85, p. 163-192.

Ragin, C.C., Becker, H.S., 1992, *What is a case? Exploring the foundations of social inquiry*, Cambridge University Press.

Raimondo E., 2023a, « Traçage de processus », Fiche méthodologique, LIEPP, n°42.

Raimondo E., 2023b, « Traçage de processus », in Revillard, A. (éd.), 2023, Méthodes et approches en EPP, Québec, Editions science et bien commun, p. 193-203.

Revillard A. (éd.), 2023, *Méthodes et approches en évaluation des politiques publiques*, Québec, Editions science et bien commun.

Riché M., 2012, «Theory Based Evaluation: A wealth of approaches and an untapped potential », Helsinki Biennal Conference of the European Evaluation Society, 4th October.

Ridde V., Coulibaly A., Gautier L., 2023, « Les études de cas », in Revillard, A. (éd.), 2023, Méthodes et approches en EPP, Québec, Editions science et bien commun, p. 181-192.

Schaffer J., 2006, « Le trou noir de la causalité », *Philosophie*, n°89, [traduction par M. Kistler et P. De Brabanter], p. 40-52.

Smith H. L., 2013, « La causalité en sociologie et démographie. Retour sur le principe de l'action humaine », Philadelphia, PA: Population Studies Center, University of Pennsylvania, PSC Working Paper Series, PSC 13-11.

Supiot A., 2015, *La gouvernance par les nombres. Cours au Collège de France (2012-2014)*, Paris, Fayard, coll. Poids et mesure du monde.

Terzi C., 2021, « Comment peut-on établir des faits par l'image », *Pragmata*, n°4, p. 16-102.

Tsing, A.L., 2021, « Vers une théorie de la non-scalabilité ». *Multitudes*, n°82, p. 65-71.

Weiss C., 2001, What Kind of Evidence in Evidence-Based Policy? In Third International, Inter-disciplinary Evidence-based policies and Indicator system Conference.

Zamora P., 2011, « La méthode d'évaluation aléatoire : apports et limites », *Tracés. Revue de Sciences humaines*, n°11, p. 175-186.