Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

L'ÉVALUATION DES POLITIQUES DE L'EMPLOI : L'ÉCONOMISTE, LE SOCIOLOGUE ET L'EXPERT

Jérôme GAUTIÉ
Centre d'économie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne

Socio-économie du travail (SET) Vous avez assuré la coordination de conseils scientifiques ou comité d'évaluation d'institutions ou dispositifs dans le domaine de la politique de l'emploi : présidence du conseil scientifique de Pôle emploi entre 2013 et 2023, présidence du comité scientifique d'évaluation de la Garantie jeunes (GJ) entre 2013-2018, et du plan « un jeune, un mentor », depuis 2023. Comment percevez-vous votre rôle ?

Jérôme GAUTIÉ (JG) L'évaluation des politiques publiques dans le cadre de ces instances implique des acteurs divers – académiques (qui eux-mêmes peuvent appartenir à des disciplines différentes : économistes sociologues, politistes, juristes, psychologues…), représentants des administrations publiques commanditaires ou parties prenantes, acteurs de terrain, notamment en charge de la mise en œuvre, représentants des bénéficiaires (en l'occurrence les jeunes dans le cadre de la GJ ou de « 1 jeune, 1 mentor »). Je viens moi-même du monde académique (en tant qu'universitaire, spécialiste d'économie du travail). Mon rôle est d'essayer de faire en sorte que les échanges entre ces différents acteurs soient les plus fructueux possibles, avec pour objectif final d'éclairer au mieux le décideur public qui commandite l'évaluation. Le processus d'évaluation dans le cadre de ces comités consiste à élaborer une « expertise » définie ici comme un avis éclairé par la contribution de l'ensemble des acteurs mobilisés. Ainsi conçue l'expertise ne se réduit pas à une expertise scientifique confiée aux académiques – même si celle-ci en est une composante importante.

SET C'est ce que vous avez fait pour la Garantie Jeune avec notamment un enjeu d'articulation entre du qualitatif et du quantitatif ?

JG Pour partir du cas de la GJ, nous avons commandité deux études qualitatives – l'une auprès des jeunes bénéficiaires[1], l'autre sur la façon dont la mesure a été mise en œuvre par les missions locales mais aussi sur comment elle transformait les pratiques de ces dernières[2] – et une étude quantitative d'impact[3]. Mais sur la question des approches, je préfère distinguer, quitte à être un peu caricatural, les études selon leur ancrage disciplinaire, économie ou sociologie, parce qu'on n'a pas seulement affaire à des différences de méthodes mais plus largement à des épistémologies différentes qui débouchent sur des apports différents en matière d'évaluation : l'évaluation d'impact pour l'économie, l'analyse du rôle des acteurs et du contexte pour la sociologie.

[1] Couronné, J., Sarfati, F., (2018), « Une jeunesse (in)visible : les 'Neets vulnérables' de la Garantie jeunes », *Travail et Emploi*, n°153, p. 15-40.
[2] Farvaque, N., Tuchszirer, C., (2018), « La Garantie jeunes en pratique dans les missions locales : une expérimentation encadrée », *Travail et Emploi*, n°153, p. 41-66.
[3] Gaini, M., Guillerm, M., Hilary, S., Valat, E., Zamora, P., (2018), « Résultats de l'évaluation quantitative de la Garantie jeunes. Quels publics, quels accompagnements et quelles trajectoires des bénéficiaires ? », *Travail et Emploi*, n°153, p. 67-88.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

SET Quelles sont les spécificités épistémologiques de l'évaluation d'impact telle que l'entend l'économie ?

JG L'évaluation d'impact passe par l'établissement de la causalité – entre un effet mesuré (par exemple en matière d'accès à l'emploi) et un « traitement » (le fait d'avoir bénéficié de la mesure étudiée). La façon d'établir la causalité est donc essentielle pour l'économiste : David Card, prix Nobel pour ses contributions dans le domaine de l'économie empirique, distingue entre l'approche fondée sur un modèle (*model based approach*) et l'approche fondée sur un protocole (*design based approach*)[4]. La première n'envisage la causalité que dans le cadre d'une théorie affirmant que « X cause Y ». L'évaluation d'impact revient alors à tester un modèle théorique et « l'étalon or » de cette démarche est de partir d'un « modèle structurel » (et non pas « sous forme réduite »). Cela prend la forme d'un système d'équations traduisant notamment les comportements microéconomiques. Principale difficulté dans cette démarche : ajuster les données au modèle théorique et estimer un système d'équations complexe. Pour la seconde approche (la « *design based approach* »), établir la causalité exige de concevoir un protocole empirique (un « *research design* »), une « manipulation », dans laquelle « X cause Y » (selon le principe « *no causation without manipulation* »). Ce protocole définit un « contrefactuel » à partir duquel on met en évidence la causalité ; « l'étalon or » de cette démarche est l'expérimentation contrôlée aléatoire (« *random controled trial* » ou RCT), ce qui vaut aux promoteurs de cette approche d'être qualifiés, non sans une pointe de critique, de « *randomistas* » par Angus Deaton. Au sein d'une même population, on tire au sort un groupe de « traitement » qui va bénéficier de la mesure évaluée, et un groupe de « contrôle » qui n'en bénéficie pas, et on regarde comment évolue la situation des deux groupes. Mais cette approche se décline aussi en « étalons » argent, bronze, etc. à travers une variété de méthodes « quasi-expérimentales ». La *design based approach* s'impose à partir du début des années 1990 à la faveur d'une « révolution de la crédibilité »[5] notamment dans le domaine de l'évaluation de politiques publiques. Elle y a acquis une très grande autorité. Cette démarche peut fournir des preuves « dures » (*hard evidence*, dans la terminologie anglo-saxonne) – selon la qualité du protocole empirique d'identification causale mis en œuvre – ou encore, selon ce que l'on nomme sa « validité interne ». Mais « l'étalon or » de cette démarche, la RCT, s'avère souvent compliquée à mettre en œuvre. En pratique, on doit souvent faire avec des formes plus ou moins dégradées de cet étalon. Cela a été par exemple le cas pour l'évaluation de la GJ. Mais se pose aussi le problème de « validité externe », c'est-à-dire la « généralisabilité » des estimations obtenues au-delà de la population et du contexte étudié dans le cadre de l'expérimentation. Comme le souligne Dani Rodrik : « *The 'hard evidence' from the randomized evaluation has to be supplemented with lots of soft evidence before it becomes usable* »[6].

SET Qu'en est-il en sociologie : comment l'évaluation de politique publique est-elle appréhendée ?

JG Les sociologues ne se mettent pas spontanément dans une posture d'évaluation… une certaine sociologie critique rejette même ouvertement cette posture, considérant la sociologie davantage comme « un sport de combat » (pour reprendre le titre d'un documentaire consacré à Pierre Bourdieu) plutôt que comme un art de l'aide à la décision

[4] Card, D., (2012), "Model-Based or Design-Based? Methodological Approaches in Empirical Micro", The University of Michigan's Woytinsky Lecture, April 27th.
[5] Angrist, J. D., Pischke, J.-S., (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspective*, 24 (2), p. 3-30.
[6] Rodrik, D., 2008, "New Development Economics: we shall experiment, but shall we learn?" Harvard Kennedy School of Government, n°RWP08-055.

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

publique. Mais au-delà de cet aspect, des différences théoriques et épistémologiques séparent les démarches d'évaluation des sociologues et des économistes. Au risque de caricaturer, le clivage fondamental tient à ce que l'économiste voit dans la réalité *un donné*, que l'on s'efforce de mesurer à l'aide de *données*, là où le sociologue voit une construction sociale. D'où l'importance accordée par ce dernier au rôle des acteurs et du contexte. Le dispositif de politique publique à évaluer et les représentations qu'en ont les agents relèvent d'une construction sociale qu'il s'agit d'analyser.

SET Qu'est-ce que cela apporte à l'exercice d'évaluation ?

JG On peut illustrer le propos avec la GJ. Il s'agissait d'un dispositif complexe où les questions relatives à la conception du dispositif et aux représentations que s'en font les acteurs se posent particulièrement. La question n'est pas seulement comment les « bénéficiaires » comprennent le dispositif et s'ils conformeront leur comportement aux attentes de ses concepteurs c'est aussi celle des représentations *a priori* des acteurs qui participent à la mise en œuvre. La GJ était par exemple catégorisée par certains acteurs du secteur social comme un dispositif de politique de l'emploi (donc par définition inadapté aux jeunes qu'eux-mêmes suivaient), avec une vision segmentée de ce qui relève de « l'emploi », et de ce qui relève du « social ». Or l'un des objectifs du dispositif était au contraire d'essayer de dépasser ce clivage, et d'aller chercher des jeunes qui n'étaient pas dans les publics cibles de la politique de l'emploi traditionnelle. Et pour cela de mobiliser les acteurs locaux (assistants sociaux, éducateurs de rue, etc.) pour qu'ils envoient ces jeunes vers le dispositif. Cela s'est heurté aux réticences de certains d'entre eux. De façon plus générale, l'analyse des acteurs renvoie à la question de ce qu'est un dispositif public, et au fait qu'il ne peut être défini indépendamment de ses conditions de mise en œuvre.

SET En quoi ?

JG En ce qu'un dispositif public est ce qu'en font les acteurs qui le mettent en œuvre au niveau local. C'est particulièrement le cas dans des dispositifs « complexes », comme la GJ, qui consistait en une articulation entre accompagnement collectif et accompagnement individuel, associée à une allocation mais soumise à une contrepartie en termes d'engagement du jeune bénéficiaire, à une nouvelle conception de rôle de l'agent de la Mission locale (la « médiation active »), au principe du « *work first* », consistant à donner la priorité aux expériences en entreprise… Pour forcer le trait, c'était aux acteurs locaux de construire la mesure – certes avec des consignes, certes de façon très cadrée par la DGEFP – à partir d'une boîte à outils qu'on mettait à leur disposition. Ce qui fait partie de l'expérimentation, c'est aussi comment est construite la mesure. Dans le détail, il y avait autant de dispositifs « GJ » que de systèmes locaux d'acteurs qui la mettaient en œuvre (avec au centre la Mission locale). Nous avons été ainsi amenés à distinguer l'« offre institutionnelle », qui correspond au cahier des charges de la DGEFP, et l'« offre organisationnelle », qui renvoie au dispositif tel qu'il a été effectivement mis en œuvre, c'est-à-dire construit et offert aux bénéficiaires au niveau local dans chaque Mission locale, mais aussi plus largement par chaque système d'acteurs locaux : les commissions locales d'attribution et de suivi composées d'acteurs très divers (représentants des services de l'Etat, protection judiciaire de l'enfance, Pôle emploi, etc.), jouaient un rôle très important. Au sein des Missions locales, on a constaté notamment des différences importantes dans l'organisation de l'accompagnement collectif. Au sein des commissions, la diversité découlait notamment des critères de sélection – selon le poids accordé à la logique « d'éligibilité » (selon laquelle un jeune a droit à la GJ à partir du moment où sa situation répond aux critères administratifs d'éligibilité), ou au contraire à une logique plus sélective, du « bien-fondé », visant à s'assurer que le dispositif était bien adapté au jeune, mais aussi

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

réciproquement, et que le jeune était suffisamment motivé pour supporter un dispositif d'accompagnement renforcé.

Au total le dispositif n'est pas donnée *ex ante*, comme quelque chose qui s'impose de façon uniforme aux individus et qu'on peut évaluer tel quel, la modalité de construction de la mesure elle-même fait partie de l'expérimentation. On dit « j'évalue la GJ » mais en fait non, on évalue un certain nombre de dispositifs en acte (pour ainsi dire) recouverts par le terme GJ, qui ont été déployés, ont été mis en œuvre de façon différente selon la mission locale, selon le contexte, etc.

SET C'est là que se joue la complémentarité entre économie et sociologie ?

JG Oui, là le sociologue est très précieux parce qu'il va voir justement comment se construisent les choses… ce qui reste souvent une boîte noire pour l'économiste ou le statisticien ! Indépendamment même de la question de l'hétérogénéité dans la mise en œuvre, l'analyse qualitative permet de comprendre ce qui se passe sur le terrain, comment fonctionne le dispositif, quels sont les mécanismes à l'œuvre qui contribuent à expliquer les effets mesurés par l'étude quantitative. Cette dernière se focalise sur la question « est-ce que ça marche ? » (et si oui, quel est l'ampleur de l'effet, pour quel coût ?) sans toujours pouvoir déceler « comment ça marche ? » et donc *in fine* : « pourquoi ça marche ? (ou pas) ».

SET Si on en revient à la diversité de mise en œuvre, ceci renvoie à « l'hétérogénéité du traitement » comme dirait l'économètre ?

JG Si la taille de son échantillon le permet, l'économètre peut repérer l'hétérogénéité dans la mise en œuvre en introduisant par exemple des indicatrices de territoires et ou Mission locales dans ses régressions. Son souci est que cette hétérogénéité ne soit pas trop importante pour que l'effet moyen mesuré soit « statistiquement significatif ». Comme noté plus haut, l'enjeu de l'étude quantitative est le plus souvent de déterminer si la mesure marche (un peu, beaucoup) ou pas, moins de comprendre (avant tout, faute de données) pourquoi elle marche mieux par exemple dans tel territoire que dans tel autre. L'étude qualitative (du moins quand elle porte sur plusieurs « terrains ») s'intéresse d'emblée à la diversité. En montée en généralité, si le nombre de cas étudiés le permet, elle peut déboucher sur des typologies – en plaisantant on pourrait opposer la secte des « tipologistas » à celle des « randomistas » pour reprendre l'expression de Deaton. Des typologies de quoi ? Par exemple d'agents locaux de l'emploi qui, en fonction de leur conception, vont accompagner différemment ou des typologies de missions locales, ou de systèmes locaux d'acteurs, etc.

SET Cela enrichit l'analyse économétrique ?

JG Oui, cela la complète, et cela peut contribuer aussi à l'analyse des conditions de la « validité externe » de l'expérimentation (ou quasi-expérimentation) quantitative (c'est-à-dire de sa « généralisabilité »), en apportant des éléments de compréhension du rôle du contexte dans le fait que ça marche plus ou moins bien. Mais cette complémentarité se joue aussi sur un autre terrain, plus en amont. L'analyse qualitative (et la démarche sociologique qui la porte) permet d'élargir les questions de l'évaluation. Au-delà des variables d'intérêt classiques, comme le retour à l'emploi, la nature de l'emploi, etc., qui sont bien sûr centrales quand on évalue un dispositif de politique « de l'emploi », d'autres variables pertinentes peuvent apparaître. Dans le cas de la GJ par exemple : l'accès à l'autonomie, le bien-être des bénéficiaires... l'enquête sociologique a permis de mettre en lumière un rôle de (re-)socialisation du dispositif – aussi bien en termes d'intégration (notamment par la logique de cohorte) que de régulation (de par l'encadrement lié à la logique de contrepartie), si on reprend les deux piliers de l'approche durkheimienne. Apporter un regard sociologique, c'est

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

aussi souvent élargir le questionnement des variables d'intérêt et permettre de dépasser « l'effet lampadaire »…

SET C'est-à-dire…

JG L'effet lampadaire renvoie à l'anecdote de la personne plongée dans le noir, qui recherche la clé qu'elle a perdue seulement sous le lampadaire parce que c'est le seul endroit où il y a de la lumière. En matière d'évaluation, on est tenté de ne s'intéresser qu'aux effets que l'on peut le mieux (ou le moins mal) mesurer, quantifier – par exemple les effets emploi pour la GJ, qui sont plus « quantifiables » que des effets sur l'autonomie ou le bien-être (cependant nous avons eu certains éléments sur ces variables grâce à une enquête de suivi). Les sociologues, par les questions qu'ils posent, sont des acteurs importants de cette négociation sur « qu'est-ce qu'il faut évaluer ? » et quelles sont les variables d'intérêt, etc. Mais ce ne sont pas les seuls : au sein du comité scientifique de la GJ, d'autres acteurs, en premier lieu les représentants d'associations de jeunes, ont insisté pour élargir les questionnements et ne pas le réduire au seuls « effets emploi ».

SET La référence à ces autres acteurs nous ramène aussi à la distinction entre l'expertise scientifique et l'expertise en un sens plus large… En quoi consiste plus précisément cette dernière ?

JG L'expertise telle que je la définis ici renvoie à l'avis motivé que l'on va donner au décideur public, commanditaire de l'évaluation, et qui va prendre concrètement la forme d'un rapport. Comme je l'ai noté plus haut, cette expertise intègre mais ne se réduit pas à l'expertise scientifique fournie par les académiques mobilisés pour mener les études quanti et quali – ce que certains d'entre eux ont parfois du mal à comprendre. Ce qui donne parfois lieu à des dérives « scientistes » – le fait de parler avec toute l'autorité que confère « la science » – avec d'autant plus de capital symbolique que l'on est reconnu dans le monde académique. Un exemple justement opposé à cette dérive dans le champ de l'évaluation est Esther Duflo qui, malgré son prix Nobel, reste toujours prudente et modeste dans ses affirmations. Ces dérives écrasent toute la richesse qui doit rester celle de ce que j'appelle « le moment de l'expertise ».

SET Où situez-vous ce moment de l'expertise ?

JG En fait à toutes les étapes du processus d'évaluation, en complémentarité avec l'expertise scientifique. D'abord, dans la définition des questions de l'évaluation, en amont : que veut-on évaluer ? Quelles variables d'intérêt ? En évitant « l'effet lampadaire ». Lors de la discussion du protocole et des résultats des études scientifiques commandités. C'est en ça que les instances type comité d'évaluation sont utiles, parce qu'elles inscrivent l'évaluation dans un processus collectif qui doit être pluraliste et pragmatique. Dans ces instances, il s'agit de construire une intelligence collective à partir d'éclairages divers (tous les acteurs présents ont un certain degré d'expertise) et d'éléments d'appréciation qui ont un statut plus ou moins « scientifique » – de l'*evidence* plus ou moins *hard* ou *soft* pour reprendre des termes anglo-saxons assez parlants. Selon moi, il s'agit de garantir en quelque sorte une continuité entre l'académique et le non académique d'où l'importance de la composition de ces instances et de leurs modalités de fonctionnement.

SET D'ailleurs, « l'évaluation pluraliste » était une référence du conseil national de l'évaluation dans les années 1990 et la variété des protagonistes étaient alors beaucoup plus large qu'aujourd'hui : par exemple il y avait notamment les partenaires sociaux. Ce n'est plus le cas dans la plupart des instances *ad hoc*…

JG Un terme aussi beaucoup usité et qui a tendu un peu à disparaître était celui de « diagnostic partagé ». C'est pour ça que j'insiste sur la question de la composition des

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

instances d'évaluation, et la production d'une intelligence collective. Et l'avis motivé que formule une telle instance, cette phase d'expertise, est toujours basée sur la mobilisation d'éclairages de différentes natures, pas seulement de productions « scientifiques » mais en fait un *continuun* entre *hard* et *soft evidence*, un faisceau d'indices pondérés par leur crédibilité/généralisabilité en évitant les effets d'autorité, mais sans tomber symétriquement dans le relativisme.

SET Et quel statut donnez-vous à cet « avis motivé » dans l'élaboration des politiques publiques ?

JG Je vais ici opposer de façon très caricaturale deux conceptions (au double sens de représentation et d'élaboration) des politiques publiques : la conception technocratique et la conception pluraliste et interactive. Ces deux conceptions concernent aussi l'évaluation. La conception technocratique repose sur une logique de planification et de contrôle (*planning and control*), selon un schéma vertical, descendant, et séquentiel : le dispositif est conçu au niveau central, par des « sachants », il s'impose ensuite au niveau local sans que les acteurs le mettant en œuvre et/ou les « bénéficiaires » qu'il vise ne soient consultés. Enfin il est évalué « scientifiquement » par d'autres « sachants », et selon le résultat on le garde (ou on le généralise) ou on l'abandonne (« stop ou encore »).

SET On reconnaît bien une tendance française mais du coup : en quoi consiste le mode « pluraliste et interactif » ?

JG Oui, la culture française est assez propice à la dérive « technocratique ». On peut lui opposer un cadre d'expérimentation entendu comme un processus interactif et intégré d'apprentissage et d'amélioration continus (*process and learnng*). L'expérimentation ne se réduit pas ici à l'expérimentation scientifique contrôlée même si celle-ci peut en être une composante. Plutôt que de s'en remettre uniquement à des « sachants », scientifiques et/ou technocrates, ce mode d'élaboration des politiques publiques va chercher à construire une intelligence collective dans une démarche pluraliste – en ce qu'elle mobilise des expertises d'origines différentes, y compris celle des acteurs de terrain – et pragmatique. Dans ce cadre, l'expérimentation est aussi un outil pour les acteurs dans une démarche d'amélioration continue. L'expérimentation sociale du dispositif « Territoires Zéro Chômeurs de Longue Durée » illustre une telle approche.

SET Si on revient au cas de la GJ, c'est donc cette seconde conception que vous vous êtes efforcé de mettre en œuvre ?

JG Oui, car c'était aussi la demande du commanditaire – en l'occurrence la DGEFP, qui avait conçu le dispositif, à l'issue d'ailleurs d'un processus lui-même interactif (dans le cadre notamment de la Commission Gurgand-Wargon) s'appuyant sur l'expérimentation d'autres dispositifs. Depuis le début des années 2000 on s'est éloigné en France du « mode technocratique » dans le domaine des politiques du travail et de l'emploi (même s'il reste assez prégnant dans certains domaines, on peut penser notamment au salaire minimum, quand on compare par exemple le système du « Comité d'experts » français à la *Low Pay Commission* au Royaume-Uni, aussi bien dans leur composition que dans leur mode de fonctionnement). Dans cette démarche, il y a une bonne articulation à trouver (aussi en termes temporels) entre les conditions de mise en œuvre d'un protocole « scientifique » d'évaluation – en croisant le quali et le quanti – et l'élaboration d'une expertise plus large. Dans une telle conception, la question n'est pas seulement de savoir si la mesure évaluer « marche » ou « ne marche pas », mais ce qui fait qu'elle marche plus ou moins bien, et ce que l'on peut en tirer comme leçon d'amélioration et d'adaptation du dispositif – c'est le processus d'apprentissage. L'évaluation ne s'adresse pas seulement au décideur public au niveau central, mais doit aussi être appropriée

Ce document est la version pre-print de l'article paru dans *Socio-économie du travail*, n°14, 2023-2. DOI : https://doi.org/10.48611/isbn.978-2-406-17798-2.p.0067.

par les acteurs de terrain qui mettent en œuvre, et/ou sont parties prenantes. Le rapport « final » de l'évaluation, qui marque en même temps la fin du Comité scientifique – pour la GJ, celui-ci a fonctionné pendant cinq ans, ce qui est une durée exceptionnelle – n'est qu'une étape si on veut véritablement un processus d'interactions (verticales et horizontales) et d'apprentissage *continus*, qui permet d'identifier les bonnes pratiques et leurs conditions d'émergence, pour pouvoir les diffuser. Pour cela, il faut maintenir des instances d'échange du genre comité de pilotage ou de suivi.