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.0055.

ÉVALUER LES POLITIQUES DE L'EMPLOI (ET DE SANTÉ) : DIX ANS DE PRATIQUE EN SERVICES STATISTIQUES MINISTÉRIELS

Mathilde GAINI
Dares, Sous-direction du suivi et de l'évaluation des politiques de l'emploi et de la formation professionnelle

Socio-économie du travail (SET) Administratrice Insee, vous avez une riche expérience de l'évaluation des politiques de l'emploi à la Dares, principalement selon des approches quantitatives, mais aussi qualitatives puisque vous avez participé aux évaluations du revenu contractualisé d'autonomie (RCA) et de la Garantie jeunes (GJ) qui articulaient volets quantitatif et qualitatif. Pour commencer, pourriez-vous nous dire quelles sont, selon vous, la portée et les limites de l'évaluation d'impact contrefactuelle ?

Mathilde GAINI (MG) Comme leur nom l'indique, les évaluations d'impact contrefactuelles ont pour objectif d'estimer l'impact d'un dispositif, par rapport à une situation contrefactuelle dans laquelle le dispositif n'aurait pas existé. Pour ce faire, regarder les trajectoires des bénéficiaires ne suffit pas, même par exemple dans le cadre d'une comparaison avant/après, parce qu'il a pu se passer de nombreuses choses sans lien avec le dispositif entre ce « avant » et ce « après », qui ont pu affecter la trajectoire des bénéficiaires. C'est pour cela que sont réalisées les évaluations contrefactuelles, que ce soit dans le cadre expérimental des évaluations aléatoires (randomisées), ce qui a été fait pour le RCA, ou qu'on se place dans une situation quasi-expérimentale comme pour la GJ… Dans ce dernier cas, l'idée est de s'approcher au maximum du cadre d'une expérimentation randomisée, en recherchant des groupes témoins constitués « par hasard » (à la faveur par exemple d'évènements fortuits ou de de seuils instaurés par les politiques publiques). Ce genre d'approche est plus facile dans le cadre d'expérimentations « locales » parce qu'avec peu de gens « traités », on a plus de chance de trouver des gens qui leur ressemblent mais qui n'ont pas bénéficié de la mesure. Mais dans le même temps, quand on évalue l'impact de dispositifs très ciblés, on a aussi le risque d'avoir des territoires ou des populations très particulières et donc, paradoxalement, d'avoir plus de difficultés à trouver de « bons » témoins. Et puis il y a aussi le risque que les résultats n'aient qu'une faible validité externe parce qu'ils portent sur des cas très particuliers.

SET Ça a été un enjeu pour l'évaluation de la GJ ?

MG Oui en effet, parce que les premiers territoires à expérimenter le dispositif, ceux de la « première vague », n'ont pas été tirés aléatoirement, même parmi ceux qui avaient candidaté pour participer à l'expérimentation. Il s'agissait principalement de territoires parmi les plus défavorisés. Il était donc nécessaire de bien contrôler du fait que ces territoires étaient différents des autres : il ne suffisait pas de juste les comparer à des territoires sans GJ. C'est pour cette raison que nous avons appliqué la méthode de double-différences : non seulement, il fallait comparer les jeunes éligibles à la GJ à d'autres

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.0055.

jeunes éligibles ailleurs, mais en plus il fallait rajouter une comparaison avec d'autres jeunes qui eux n'étaient pas éligibles, à la fois dans les territoires de la GJ et dans les autres territoires, pour pouvoir contrôler d'évolutions différentes entre les territoires au cours du temps qu'on aurait pu à tort attribuer à un effet du dispositif. Une difficulté à laquelle nous avons été confrontés, et qu'on retrouve souvent dans ces évaluations contrefactuelles quasi-expérimentales, c'est que la plupart du temps on ne peut pas se contenter de comparer les seuls bénéficiaires de la politique à d'autres jeunes éligibles mais qui n'en ont pas bénéficié.

SET Pourquoi ?

MG Parce que souvent, le « take-up », c'est-à-dire le taux de recours à la politique, n'est pas de 100 %, et que le fait de recourir à la politique n'est potentiellement pas aléatoire. Il est donc possible qu'il y ait un biais de sélection, c'est-à-dire que les éligibles bénéficiaires soient différents des éligibles non-bénéficiaires. Dans le cas de la GJ, nous avons donc reconstitué des groupes d'éligibles dans les territoires traités comme dans des territoires témoins. Pour cela, nous avons mis en place un dispositif particulier de recueil de données. On dispose en effet la plupart du temps de données sur les bénéficiaires de politiques publiques, notamment car un certain nombre d'informations sont renseignées lorsqu'un bénéficiaire entre en dispositif. En revanche c'est souvent plus difficile d'avoir de la donnée sur les éligibles. Ce, pour deux raisons principales : soit parce que les critères retenus concernent des informations auxquelles nous n'avons pas accès, soit parce que les critères ne sont pas « stricts » et qu'il y a une marge d'appréciation, par exemple du conseiller du service public de l'emploi. Dans le cas de la GJ, l'éligibilité était assez clairement définie – il s'agissait des jeunes en situation de précarité – mais reposait sur des informations qui ne sont pas enregistrées en routine par les missions locales dans leur système d'information (par exemple, les revenus). Nous avons donc mis en place une enquête *ad hoc* auprès des conseillers de mission locale, dans laquelle nous leur avons demandé de renseigner pour tous les jeunes qu'ils rencontraient s'ils répondaient d'après eux aux critères d'éligibilité de la GJ. Nous leur avons aussi demandé un certain nombre d'informations pour décrire cette population.

SET C'est aussi ce que vous avez fait avec l'enquête emplois d'avenir ?

MG Non, pour évaluer les emplois d'avenir, nous avons privilégié une évaluation par *matching*, donc avec une force de preuve plus faible. Avec cette méthode, il faut disposer d'un maximum d'informations possible sur les individus, afin de contrôler du biais de sélection sur les bénéficiaires. Nous avons donc mis en place une enquête auprès de bénéficiaires et d'une population témoin ayant eu des contacts avec la mission locale, et nous avons posé de nombreuses questions aux jeunes sur leur passé (parcours en emploi, scolarité, situation familiale…) pour essayer de trouver le groupe témoin le plus comparable possible. Au-delà de données riches pour définir le groupe témoin, nous avons également besoin de données riches pour analyser l'impact des dispositifs. Et pour aller au-delà de l'effet lampadaire…

SET C'est-à-dire ?

MG Aujourd'hui, avec la déclaration sociale nominative (DSN) il est plus aisé que par le passé de documenter le parcours en emploi des personnes, mais, idéalement, ce qu'on souhaite couvrir, c'est un spectre de variables de résultat (*outcomes*) plus large que le seul accès à l'emploi. Par exemple, dans le cas de la GJ, nous nous sommes intéressés à l'accès à l'autonomie, au permis de conduire, au logement, au renoncement au soin mais aussi à des variables subjectives : bien-être, confiance en l'avenir… Or, pour

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.0055.

la plupart de ces données, il est nécessaire de recourir à des enquêtes. D'ailleurs l'intérêt des données administratives telles que la DSN, c'est aussi de libérer de la place dans nos enquêtes pour collecter des informations qu'on ne peut pas obtenir autrement qu'en interrogeant les personnes, notamment des questions subjectives, mais pas que. Mais produire des enquêtes a aussi un coût, de conception, de collecte, et pour les enquêtés.

SET Ça nous amène à un autre sujet important : la structuration de la démarche d'évaluation en présence de ressources contraintes.

MG C'est en effet un sujet important. De plus en plus d'évaluations sont prévues dès la création de dispositifs, ce qui est une avancée. Mais ça vient avec une difficulté. Le revers de la médaille, c'est qu'il faut réussir à suivre cette masse d'évaluations potentielles : quels moyens met-on en face pour assurer ces évaluations, qui sont coûteuses ? On mobilise du temps humain, on rédige des appels à projet de recherche et on lance des marchés pour avoir des évaluateurs, on conçoit et on finance des enquêtes... Cela pose la question « où dois-je mettre mon effort d'évaluation ? » Illustration très concrète dans mon quotidien lorsque j'étais au Ministère de la Santé : évaluer systématiquement plus de cent projets, comment dire ? C'est un peu titanesque.

SET Oui parce qu'il faut dire qu'entre 2018 et 2023, vous étiez en poste à la Drees, qui est en charge avec la Caisse nationale d'assurance maladie (Cnam) de l'évaluation des expérimentations dites « article 51 » dans le domaine santé-solidarité. À ce moment-là votre activité d'évaluation est devenue industrielle pour ainsi dire…

MG En effet, la Loi de Finance de la Sécurité Sociale de 2018[1] (plus précisément son article 51) a ouvert la possibilité de mener des expérimentations dites « d'innovation organisationnelle en Santé » pouvant déroger au droit commun, notamment sur le financement. Par exemple en intégrant des forfaits à la place de tarification à l'activité, ou en remboursant certains soins, tels que la consultation de psychologues, qui ne l'étaient pas… l'idée était de faire remonter des projets du terrain, et de tester leur faisabilité et leur passage à l'échelle. Aujourd'hui, plus de 140 projets ont été autorisés…

SET Et vous avez dû définir un cadre commun d'évaluation pour tout ça ?

MG Précisément : la loi prévoit l'évaluation de chaque expérimentation. Il a donc fallu penser un cadre conceptuel commun, dans le contexte d'une évaluation qui se voulait pragmatique. Le principe a été d'impliquer systématiquement les porteurs de projet. D'abord en organisant sur site leur rencontre avec l'équipe chargée de l'évaluation de leur projet, puis en leur permettant de relire tous les rapports rédigés dans le cadre de l'évaluation : rapports intermédiaires et rapport final. Sur cette base, l'enjeu était d'orchestrer des évaluations articulant trois dimensions : un questionnement sur la faisabilité, mené plutôt selon une démarche qualitative, une mesure d'efficacité/efficience croisant qualitatif et quantitatif et une réflexion sur la reproductibilité/généralisation adossée à de l'enquête qualitative.

SET Alors justement, comment ça s'est organisé ?

MG Au début, nous n'avions pas forcément anticipé qu'il y aurait autant de projets à évaluer. Nous avons donc pensé un dispositif d'évaluation très complet, avec du qualitatif, du quantitatif, et l'idée qu'il fallait élaborer la stratégie d'évaluation des projets sur les trois axes retenus – faisabilité, efficacité / efficience, passage à l'échelle /

[1] Loi n° 2017-1836 du 30 décembre 2017 de financement de la sécurité sociale pour 2018.

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.0055.

reproductibilité – dès le lancement de l'expérimentation… Mais avec la croissance très forte du nombre de projets autorisés, nous nous sommes retrouvés face à la difficulté de ne pas forcément arriver à tout faire, et d'ailleurs ce n'était pas forcément pertinent. En effet, nous nous sommes rendus compte assez vite qu'il y avait un écart entre ce qui avait été pensé par les porteurs et ce qui se réalisait vraiment sur le terrain. À ce moment-là, on s'interroge sur la façon la plus optimale de répartir notre effort, et notamment de « détecter » les projets les plus intéressants, ou les plus transformatifs. Il s'avère qu'en fait, c'est assez difficile de savoir en amont ce qui va être intéressant, ce qui va fonctionner…

SET Quel choix a été fait ?

MG Nous avons écarté l'idée de définir *ex ante* un plan d'évaluation complet, notamment parce qu'il était trop tôt pour juger de la pertinence de mobiliser des analyses quantitatives avancées. Pour les rapports intermédiaires, nous avons donc décidé de produire systématiquement des analyses qualitatives afin de décrire la faisabilité du projet et de documenter concrètement comment se faisait son déploiement sur le terrain. La décision de mobiliser des données quantitatives avancées pour les rapports finaux, notamment celles du système national des données de santé (SNDS, données de remboursement et d'hospitalisation) ne se prend que dans un second temps. Cependant, cela ne peut pas non plus intervenir trop tard, car accéder aux données prend du temps (autorisations d'accès, mise à disposition), comme mettre en place un recueil de données *ad hoc* si nécessaire.

SET Ça suggère un séquencement dans la mobilisation des approches quali et quanti…

MG Cette complémentarité quali-quanti, on voit bien en effet dans ce cas qu'elle s'articule dans le temps. Parce que ces deux types d'évaluation répondent à des questions différentes. En premier lieu, cela me semble essentiel de commencer par du quali pour décrire ce qu'il se passe et analyser la faisabilité d'un projet. Ensuite on peut avoir un couplage quali-quanti, avec du quanti plutôt sur l'efficacité, l'efficience et du quali qui nous permette de comprendre quels sont les mécanismes sous-jacents à l'œuvre. Typiquement, le quanti permet le plus souvent de connaître l'impact « agrégé » d'un dispositif, alors qu'une politique est souvent complexe. Par exemple, une des expérimentations avait pour objectif d'inciter à la prescription de biosimilaires, et pour ce faire elle comportait à la fois une hausse du taux d'incitation à la prescription mais aussi le fait que contrairement au droit commun où l'incitation allait à l'hôpital, là elle allait directement au(x) service(s) où se faisaient les prescriptions. Donc une fois l'effet de l'expérimentation évalué, on ne sait pas bien par quel(s) canal(aux) il transite. C'était à ce titre intéressant d'avoir des analyses qualitatives qui nous disaient par exemple « vous savez sur le terrain, les professionnels ne savent pas forcément que l'incitation est différente du droit commun ». Mais cela n'empêche pas qu'il est très important d'avoir des évaluations d'impact contrefactuelles. J'en ai eu très fortement conscience lors de l'une des premières évaluations que j'ai faites, qui concernait le RCA.

SET En quoi ?

MG Sur le RCA, j'ai souvenir de l'évaluation qualitative qui semblait indiquer : « les conseillers trouvent ça génial, les jeunes sont beaucoup plus autonomes, ils passent leur permis, ils sont motivés pour trouver un emploi… ». Or, via l'évaluation quantitative, nous n'avons trouvé aucun effet du RCA, et pourtant nous avons regardé de nombreuses dimensions (*outcome*) grâce à des données d'enquêtes riches. D'après moi,

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.0055.

le quali ne doit pas servir à statuer sur l'efficacité d'un dispositif, parce que c'est très difficile pour les personnes interrogées de se placer dans cette situation contrefactuelle. Par contre, les personnes, et par conséquent les travaux qualitatifs, peuvent décrire précisément ce qu'il se passe et cela permet d'éclairer les résultats d'évaluations quantitatives par ce que les chercheurs observent sur le terrain.

SET Une autre dimension de l'évaluation dont nous n'avons pas encore parlé est l'analyse coût/bénéfice…

MG C'est vrai qu'une fois qu'on a réalisé une évaluation contrefactuelle qui démontre un impact positif, et qu'on a identifié les mécanismes à l'œuvre, se pose la question de l'efficience. Par exemple, toujours sur l'expérimentation biosimilaire, nous avons trouvé que la prescription de médicaments biosimilaires a nettement augmenté dans les établissements de santé traités. À première vue, c'est donc un dispositif qui fonctionne. Sauf qu'ensuite, nous nous sommes intéressés à l'efficience, et là le constat est plus nuancé.

SET Oui parce que dans l'évaluation coût-bénéfice, l'idée est d'être au plus près des préoccupations des décideurs.

MG En effet, du point de vue du décideur public, la question de l'efficacité n'est pas suffisante, on a besoin de regarder l'efficience : est-ce que les coûts engendrés par le dispositif sont a minima compensés par les gains qu'on en tire ? Or, c'est encore plus compliqué que l'évaluation contrefactuelle, pour plusieurs raisons. D'abord, il faut pouvoir rendre compte de tous les effets d'un dispositif, et comme discuté précédemment on dispose plus facilement de données sur certaines dimensions que sur d'autres. Ensuite il faut pouvoir transcrire ces effets en dépenses ou économies supplémentaires, ce qui n'est là encore pas chose aisée. Par exemple, un dispositif qui permet d'améliorer l'accès aux soins va probablement avoir un effet négatif sur les finances publiques à court-terme mais positif à long-terme car cela va éviter des prises en charge plus tardives et plus coûteuses…

SET Pourrait-on revenir à la question de la place de l'évaluation par rapport au temps politique : que tirez-vous de votre expérience en la matière ?

MG Mon expérience, c'est que c'est très rare que les politiques attendent les résultats finaux des évaluations, encore plus pour les évaluations quantitatives qui mettent souvent plus de temps que celles qualitatives à être finalisées. Parce que ça prend du temps d'avoir l'enquête réalisée ou des données administratives qui remontent, des données consolidées et des bases propres, des résultats robustes. Ça met du temps aussi de mesurer des effets post-dispositif, leur persistance. Pour la GJ, on voulait regarder plus de 18 mois après l'entrée en dispositif ce qu'il en était. Mais tout ça, ça montre en miroir l'importance de l'évaluation *in itinere*…

SET Pourquoi ?

MG L'évaluation qualitative, la première chose et le premier mérite qu'elle a pour moi, c'est qu'elle va sur le terrain et qu'elle décrit la réalité de l'expérimentation. Elle fait déjà un travail documentaire avant même de livrer des analyses sur les jeux d'acteurs, les mécanismes sous-jacents à l'œuvre, etc. Ce sont les évaluateurs quali qui vont vraiment sur le terrain et qui voient ce qu'il s'y passe. L'administration, elle pense le dispositif mais elle n'est pas forcément sur place pour faire ces observations, et ne dispose pas non plus des outils d'analyse mobilisés par les chercheurs. Or il existe le plus souvent un écart entre le dispositif tel qu'il a été conçu et sa réalité sur le terrain, et cela

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.0055.

ne provient pas uniquement du fait qu'on a des politiques nationales qui donnent un cadre et qui se déclinent ensuite dans des territoires avec des spécificités propres. Pour revenir à ce qu'on a observé avec l'article 51, on s'est rendu compte que bien que le porteur de projet échange avec l'équipe nationale « Article 51 » sur le cahier des charges, et qu'il porte lui-même le projet sur le terrain, cela n'empêche pas qu'entre ce qui a été pensé au départ et ce qui se met vraiment en place, il peut y avoir parfois de grosses différences. Et c'est déjà très important de documenter ces écarts et d'expliquer pourquoi ils existent.

SET Et le côté « *in itinere* » ? L'évaluation « au-delà de l'impact » ?

MG Avoir une évaluation *in itinere* permet de nourrir les réflexions pour potentiellement réorienter les dispositifs ou tester d'autres modalités, et disposer d'éléments, même parcellaires, qui permettent d'alimenter si besoin des décisions politiques. Mais il ne faut pas pour autant renoncer à finaliser les évaluations même après les généralisations ou les abandons de dispositifs, parce qu'il est important de construire et d'archiver les savoirs, de les capitaliser, et pour ce faire il faut qu'il y ait des écrits. Pour donner un exemple, au moment où ont été conçus les emplois d'avenir, je me souviens d'avoir recherché des analyses sur les emplois jeunes. Or, je ne sais pas si ça a été peu évalué ou si les travaux n'ont pas été conservés et valorisés, mais *de facto* nous ne disposions pas de beaucoup d'éléments. L'enjeu, c'est d'accepter qu'on continue de consacrer des personnels et des financements sur des politiques qui ne sont plus prioritaires.

SET L'utilité de l'évaluation serait plus ou moins indépendante des objectifs politiques ?

MG Elle peut quand même servir à la décision. Dans tous les cas, l'évaluation peut être porteuse de connaissances et de transformations quelle que soit l'issue de l'expérimentation par, justement, le travail réflexif qu'elle engendre et la capitalisation sur ses enseignements. Porteuse de connaissances parce qu'il y a toujours ce travail documentaire avant même l'évaluation à proprement parler. Porteuse de transformation, quand elle n'est pas faite en chambre, lorsqu'elle revient vers les porteurs de projet et pas restituée uniquement à la fin. Mais ce côté *in itinere* implique des tensions parce qu'il faut arriver à préserver l'indépendance de l'évaluation, assurer les conditions de sa qualité, etc. Et ce n'est pas toujours évident de communiquer sur une évaluation en cours, d'échanger et d'éviter les conclusions hâtives. Et puis inversement, pour l'évaluateur ça implique aussi d'accepter que l'objet évalué n'est pas totalement pur et qu'il faut quand même arriver à produire une évaluation. Impur, ça ne veut pas dire complètement brouillon : il faut quand même qu'on arrive à dire des choses. La question est « où place-t-on le curseur ? » pour pouvoir dire des choses valides mêmes si tous les canons de l'évaluation académique n'y sont pas.

SET Pourriez-vous développer sur cet enjeu de capitalisation de ce que produisent les travaux d'évaluation ?

MG Le travail réflexif et la capitalisation, ce sont des aspects essentiels. C'est toute l'importance des lieux d'échange sur les résultats des évaluations. C'est typiquement le rôle qu'a eu le conseil scientifique de l'expérimentation de la GJ. Un comité avec des personnes qualifiées qui sachent faire dialoguer les différents types d'évaluation et inscrire ces dernières dans des problématiques plus globales. L'administration a, il me semble, besoin d'acteurs externes qui souhaitent s'inscrire dans l'objectif de contribuer à la pertinence et à l'efficacité des dispositifs. Mais il faut que ces acteurs partagent aussi l'objectif d'aider à la décision ou d'améliorer la politique plus que

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.0055.

de produire une évaluation valorisable académiquement. Parce qu'effectivement, ces évaluations, ce sont souvent des objets non ou mal identifiés et reconnus dans le monde académique. Il est donc nécessaire d'entrer dans une démarche constructive, qui accepte aussi les limites de l'évaluation. Un tel positionnement peut rebuter les « puristes ». On a d'ailleurs assez peu de candidats qui acceptent de se prêter à cet exercice et le font de façon convaincante. Car l'enjeu n'est pas de produire des évaluations de mauvaise qualité, mais de tirer tous les enseignements possibles tout en ayant parfaitement conscience des limites des travaux produits.

SET Diriez-vous que le travail réalisé par les conseils scientifiques est insuffisamment valorisé ?

MG A minima, je pense que la plus-value d'avoir un conseil scientifique n'est pas toujours perçue. En particulier, un conseil scientifique, avec un président, assure normalement la production d'écrits qui rendent compte des échanges au-delà des seuls rapports d'évaluation, renforce l'appropriation de ces évaluations et les fasse dialoguer. Cela m'est apparu de façon assez claire lorsque je me suis occupée de l'évaluation autour de l'article 51 pour laquelle nous ne disposions que des rapports qualitatifs, et éventuellement quantitatifs, accompagnés d'une page de synthèse. En fait, un très grand nombre de savoirs « gris » ont été amassés collectivement, des savoirs très utiles pour alimenter les réflexions opérationnelles mais pas toujours capitalisés, car cela prend du temps et qu'il faut savoir le faire. Nous avons essayé de mobiliser des chercheurs pour travailler sur des méta-analyses, reprendre les différents rapports, essayer de les lire sous différents angles, mais c'est difficile de trouver des personnes qui soient intéressées et aient le temps de mener cet exercice.

SET Pour revenir sur le lien de l'évaluation à la décision politique, comment analysez-vous l'étape de la décision, dont vous nous disiez qu'elle intervenait souvent avant que l'évaluation ait livré toutes ses conclusions ?

MG Paradoxalement, la généralisation avant les résultats, cela crée aussi moins de pression sur les résultats des évaluations (mais cela reste une mauvaise pratique !). Lorsque la décision est suspendue à l'évaluation, cela suscite de nombreux enjeux opérationnels. C'est le cas dans le cadre de l'article 51 où il est clairement écrit que c'est sur la base des évaluations que sont prises les décisions de généralisation ou d'abandon. Dans ce cadre-là, il y a une vraie pression sur les délais de production des évaluations. Se pose aussi la question de ce que l'on fait en attendant les résultats. La loi a prévu que les résultats soient disponibles six mois après la fin des expérimentations. Mais dans les faits, on ne peut pas dire aux porteurs de projet : « attendez, on s'arrête tous parce que dans six mois, on va vous dire si on continue ou si on revient au droit commun ». Nous avons donc essayé d'avancer la production du rapport final, avec un rapport trois mois avant la fin de l'expérimentation… mais cela ne règle pas tous les problèmes. En effet, les évaluations quantitatives portent par la force des choses sur des données passées, et donc le porteur de projet peut toujours arguer du fait que les résultats ont changé sur les derniers mois et qu'il faudrait refaire l'évaluation pour en tenir compte…

SET Pas très facile en effet…

MG Et puis il y a aussi la question : que tire-t-on des résultats ? S'il est facile de prendre une décision opérationnelle lorsque les résultats sont très tranchés, que ce soit en faveur ou en défaveur d'un projet, c'est beaucoup plus difficile lorsque les rapports d'évaluation sont plus nuancés sur un dispositif, avec des axes de réussite mais aussi un certain nombre de points d'attention. Dans ce cas, plutôt que de généraliser, on

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.0055.

peut souhaiter repartir dans une phase d'expérimentation. On peut par exemple tester sur de nouveaux territoires des projets qui ont l'air d'avoir bien fonctionné mais pour lesquels on a des doutes sur la reproductibilité parce qu'ils ont été déployé dans des conditions très particulières. On peut aussi choisir de tester des cahiers des charges modifiés, quand on a des expérimentations sur des sujets proches et qu'on se dit qu'on pourrait sûrement avoir quelque chose qui soit un peu hybride et qui capitalise sur les points forts de différents projets. La difficulté de ces expérimentations « en chaine », c'est qu'il y a bien un moment où il faut s'arrêter d'expérimenter parce qu'*in fine* on souhaite plutôt proposer un dispositif pour une problématique donnée. Pour l'administration c'est difficile d'être dans un état d'expérimentation permanente : il y a la question de l'égalité de traitement sur le territoire, la complexité à gérer le hors droit commun, etc. Donc, il y a bien un moment quand même, où il faut prendre des décisions opérationnelles. Ce travail-là va au-delà de tirer les conclusions de l'évaluation, c'est de se dire « concrètement, qu'est-ce que je fais opérationnellement avec ce dont je dispose aujourd'hui comme connaissances ? ». C'est aussi un travail complexe pour lequel on a besoin de passerelles entre évaluation et opérationnel.