

Les démarches d'évaluation en question

L'évaluation est de plus en plus souvent intégrée à la mise en oeuvre des politiques publiques. Aujourd'hui plus complexes, les démarches évaluatives gagnent en précision mais sont plus exigeantes et demandent un certain nombre de précautions d'usage. L'ensemble des acteurs doit néanmoins se convaincre de leur intérêt, dans l'objectif d'accroître l'efficacité des mesures.

emploi
évaluation
insertion
formation
certification
carrière
travail méthodologie

Pascale Rouaud
Emmanuel Sulzer
(Céreq)

L'évaluation est aujourd'hui sur le devant de la scène et se diffuse progressivement aux différents niveaux de l'appareil administratif. Elle est désormais pensée comme une composante centrale de l'élaboration, puis de la mise en oeuvre, des politiques publiques, à la recherche d'une plus grande efficacité. C'est l'esprit même de la loi organique relative aux lois de finances (LOLF) de 2001.

Pensée ainsi, et contrairement à ce que l'on observe dans la plupart des pays anglo-saxons, la démarche évaluative commence seulement à se généraliser en France. Les panels de suivi des bénéficiaires des politiques d'emploi mis en place par la Direction de l'animation, de la recherche, des études et des statistiques (Dares) au début des années quatre-vingt dix sont ainsi un des rares exemples français d'outillage d'une démarche évaluative des politiques d'emploi par des sources statistiques longitudinales.

Dans la lignée de ces évolutions, les traditionnelles évaluations *ex ante* coexistent

de plus en plus avec des protocoles *ex post*, nourris du développement des approches longitudinales, et plus récemment en France, avec des expériences « randomisées ». Si ces protocoles, de par leurs qualités techniques, peuvent être parfois présentés comme une forme « pure » ou « aboutie » de l'évaluation, leur mise en oeuvre concrète soulève un certain nombre de questions et de difficultés.

En mai 2010 se tenaient à Marseille les XVII^{es} journées d'étude du longitudinal, organisées par le Céreq et consacrées à l'évaluation. Plusieurs enseignements ressortent des contributions à ces journées. Dans le cas de l'évaluation randomisée, plus encore que dans l'évaluation classique, certaines conditions doivent être respectées. Configurer le dispositif en amont, le compléter et l'enrichir par des approches qualitatives, concilier neutralité et association des acteurs aux différentes phases du processus apparaissent ainsi comme les éléments clés d'une évaluation réussie. ●●●

Les questions de formation, d'emploi, d'insertion professionnelle sont au coeur des débats de la société française. Depuis 40 ans, le Céreq étudie, évalue, propose ses résultats à l'ensemble des acteurs et des décideurs. Les questions relatives aux jeunes, au système éducatif, à la formation tout au long de la vie ou à l'apprentissage sont de plus en plus nombreuses et complexes.

Ce numéro est l'occasion de vous présenter la nouvelle maquette du Bref. Elle va de pair avec une déclinaison en quatre nouvelles séries :

- le Bref **Études** présente de façon concise les résultats d'une étude ;
- le Bref **Opinion** affirme un point de vue d'expert sur un débat d'actualité ;
- le Bref **Territoires** offre un angle d'attaque plus local de la part de nos centres de recherche associés en région ;
- le Bref **Ouvertures** propose un point sur les principales publications, postures et controverses.

Ces séries viendront progressivement s'intercaler au sein de la collection pour vous offrir plus de variété.

Je vous donne également rendez-vous sur le nouveau site, plus convivial et mieux adapté aux attentes actuelles des usagers.

N'hésitez pas à nous faire part de vos réactions.

Frédéric Wacheux, directeur du Céreq

Le Fonds d'expérimentations pour la jeunesse

Le Fonds d'expérimentations pour la jeunesse a été créé en décembre 2008, à l'initiative de Martin Hirsch, alors Haut Commissaire à la jeunesse, dans le but d'agir sur l'insertion sociale et professionnelle des jeunes de seize à vingt-cinq ans. Il est maintenant géré par le ministère de l'Éducation nationale, de la Jeunesse et de la Vie associative et dispose d'un budget de 250 millions d'euros.

Il a été pensé comme un laboratoire destiné à fonder les politiques publiques. Son mode d'action part d'un double constat : les initiatives locales peinent souvent à se généraliser faute d'apporter la preuve de leur efficacité ou de leur caractère reproductible, tandis que de nombreux dispositifs nationaux n'atteignent pas leurs objectifs car ils n'ont pas été testés préalablement.

Le Fonds propose donc de contribuer au financement de dispositifs innovants de politique sociale initiés à une échelle limitée et mis en œuvre dans des conditions qui doivent permettre une évaluation rigoureuse et scientifique des effets, afin de juger *in fine* de l'intérêt d'une éventuelle généralisation. Chaque expérimentation est évaluée par un organisme externe (centre de recherche, équipe universitaire ou cabinet privé).

Pour orienter les expérimentations, le Fonds réalise des appels à projets sur les thématiques qu'il juge prioritaires : lutte contre le décrochage scolaire, amélioration de l'insertion professionnelle des jeunes, lutte contre les discriminations, etc. Dans ses choix, il encourage par ailleurs les évaluations contrôlées, avec assignation aléatoire des bénéficiaires (sur le modèle des expérimentations médicales).

Pour construire ses appels à projets et sélectionner les expérimentations, le Fonds s'est doté d'un conseil scientifique dont les membres interviennent dans les jurys de sélection, aux côtés des services déconcentrés de l'État, d'acteurs de terrain et de professionnels des domaines considérés.

À ce jour, plus de 300 expérimentations ont été lancées. Le Céreq et ses centres associés, évaluateurs d'une trentaine de projets, sont fortement impliqués.

Il est bien sûr trop tôt pour dresser un bilan de cette démarche mais elle apparaît d'ores et déjà novatrice, notamment dans son souci d'intégrer les impératifs d'une évaluation *ex post* en amont de la mise en place du dispositif.

Pour en savoir plus : www.experimentationsociale.fr

●●● Tradition et innovation en évaluation

Jusqu'à présent, l'administration française favorisait principalement l'évaluation *ex-ante* ainsi que des études évaluatives *ex post* légères (cf. encadré page suivante). L'évaluation *ex ante* est avant tout prospective. Elle consiste à étudier la faisabilité et à prévoir l'impact d'un programme. L'évaluation *ex post* apprécie les effets d'une politique après sa mise en œuvre. Elle permet de comparer les résultats aux moyens engagés (mesure de son efficacité) et aux objectifs initialement fixés (mesure de son efficacité). Elle peut prendre des formes légères, via des études d'impact, des enquêtes de satisfaction ou un suivi des coûts, ou des formes plus lourdes à travers des évaluations scientifiques ou institutionnelles. Les secondes ont la particularité d'être menées collégalement par des instances *ad hoc* (praticiens, experts, scientifiques, représentants des catégories concernées) pour apprécier l'ensemble des effets de la politique. Les évaluations scientifiques lourdes sont réalisées par des

économistes et des statisticiens ; elles s'appuient sur des données individuelles et font intervenir des modèles microéconomiques pour comparer la population des bénéficiaires à une population de référence ayant des caractéristiques proches.

Une nouvelle démarche s'est développée aux États-Unis et dans certains pays européens : l'évaluation contrôlée, qui utilise la méthode « des essais randomisés » issue des protocoles médicaux. Schématiquement, la méthode consiste à constituer par tirage au sort deux groupes au sein d'une population éligible à une mesure : le premier est composé d'individus qui bénéficient du dispositif (le groupe « test »), le second d'individus qui n'en bénéficient pas (le groupe « de contrôle »). On compare ensuite la situation des individus ayant bénéficié du dispositif à celle des autres. Cette méthode garantit que les différences observées *a posteriori* sont imputables au dispositif, et non à d'éventuelles différences entre les caractéristiques individuelles des bénéficiaires et celles des non-bénéficiaires. Il est donc possible de conclure sur les effets réels de la mesure.

L'évaluation contrôlée a récemment fait son apparition en France, notamment dans le cadre des politiques sociales et du Fonds d'expérimentations pour la jeunesse (cf. encadré ci-contre). Elle est en effet bien adaptée à l'évaluation des dispositifs innovants de politique sociale réalisés dans un premier temps à petite échelle.

Une démarche qui se prépare en amont

L'évaluation a pour objectif, *in fine*, de se rendre utile à la décision. Pour cela, décrire la situation des bénéficiaires après leur passage par un dispositif ne suffit pas. Encore faut-il connaître le contexte initial et les acteurs en présence, et comprendre les processus mis en œuvre. Toute évaluation gagne donc à être pensée en amont de l'action et à observer, voire accompagner, son déroulement.

Définir les objectifs de l'évaluation avant le démarrage du projet permet d'établir un protocole qui soit compatible avec l'action et les réalités concrètes de terrain. Les indicateurs pertinents de l'efficacité d'un dispositif devront dans certains cas être construits *ex nihilo* ; c'est pourquoi il importe préalablement de définir précisément ce sur quoi devra porter l'évaluation. Dans le cas de l'innovation sociale, ce travail de définition en amont est souvent primordial, et peut nécessiter dans un premier temps de faire converger les acteurs engagés dans le projet à évaluer.

L'exemple du « décrochage scolaire », objet aujourd'hui de nombreux travaux, illustre bien cet aspect. Qui est considéré comme « décrocheur » ?

Tous les acteurs de la prise en charge du décrochage utilisent le même terme, mais chacun a sa propre définition, qui correspond à sa propre visée opérationnelle. Pour l'institution Éducation nationale, ce terme fait référence au jeune qui abandonne ses études en cours de formation. Pour la Mission locale, il s'agit du jeune non qualifié ayant déclaré une date de fin de scolarité lors de l'entretien avec un conseiller. Enfin, l'enseignant nomme décrocheur un élève qui présente en cours d'année des absences non justifiées et répétées. Avant de rédiger le protocole d'étude, il faut donc chercher à créer une culture et un langage communs.

Si le choix des indicateurs de résultats est primordial, il faut également veiller à ce que ceux-ci puissent être renseignés de manière satisfaisante. En effet, les systèmes d'information préexistants ne fournissent pas toujours les indicateurs souhaités. Dans certains cas, la procédure d'évaluation doit élaborer un système *ad hoc* de recueil de données. Dans le cas de l'évaluation contrôlée, jugée aujourd'hui pertinente pour les dispositifs innovants, c'est l'intégralité du protocole (mode de tirage, choix des populations test et contrôle, etc.) qui doit être anticipée afin de garantir la fiabilité de la démarche.

Pour des approches complémentaires

Combiner les approches quantitatives et qualitatives permet d'enrichir l'évaluation. Très schématiquement, le volet quantitatif a pour

objectif de mesurer des effets sur la population ciblée par le dispositif : est-ce que le dispositif est efficace au regard de l'objectif poursuivi ? Le volet qualitatif cherche pour sa part à décrire et comprendre pourquoi, et comment, cela marche ou pas.

L'évaluation contrôlée, fiable scientifiquement, soulève cependant plusieurs types de réticences. La première est d'ordre éthique. En effet, le tirage aléatoire implique que certaines personnes, qui auraient légitimement pu bénéficier de la mesure évaluée, n'en bénéficieront pas. La seconde est d'ordre pratique, et porte sur la faisabilité de l'évaluation : pour pouvoir être mise en œuvre, cette méthode nécessite des effectifs suffisants. Enfin, lorsqu'une action s'adresse à des groupes (classes d'étudiants, villages...), le tirage aléatoire au niveau individuel n'est pas envisageable. Des méthodes statistiques permettent néanmoins de contourner cet écueil.

La généralisation des évaluations contrôlées présente aussi des limites. Ces méthodes supposent tout d'abord une stabilité du contexte entre le moment de l'expérimentation et celui de sa généralisation. Lorsque ce n'est pas le cas, les résultats peuvent être différents lors du passage en grandeur réelle. Multiplier les évaluations dans des environnements différents peut alors constituer une parade. Autre limite, l'effet positif constaté à petite échelle lors du test peut être atténué sur l'ensemble de la population lors de la généralisation si, par exemple, le public cible est moins bien informé que lors de l'expérimentation.

Les démarches évaluatives, quelques repères

- **Évaluation ex ante** : elle est réalisée avant la mise en œuvre de l'action ou du dispositif. Elle vise à établir un diagnostic du contexte, étudier la faisabilité de l'intervention ou réaliser des scénarios d'impact prévisionnel. Elle fait intervenir des analyses d'efficacité par simulations macroéconomique, qui reposent sur des estimations chiffrées des mécanismes économiques.
- **Évaluation embarquée (ou « on-going »)** : elle se déroule pendant toute la période de mise en œuvre de l'intervention. Elle permet à l'évaluateur de comprendre le contexte et les conditions de mise en œuvre de l'intervention. Elle implique l'évaluateur et l'expérimentateur dans une démarche d'échange et favorise ainsi leur collaboration.
- **Évaluation ex post** : elle est réalisée après la mise en œuvre de l'action, à un moment où ses effets ont eu le temps de se matérialiser. Elle met généralement en œuvre des moyens importants, en s'appuyant sur des bases de données longitudinales et en mobilisant des techniques microéconomiques
- **Quasi-expérimentation** : il s'agit d'une évaluation *a posteriori*, qui consiste à comparer une population de bénéficiaires d'un dispositif à une population de non-bénéficiaires. Les études faisant intervenir un traitement économétrique appartiennent à cette catégorie, ainsi que les études de type « avant-après » ou « ici-ailleurs ». Dans une étude de type « ici-ailleurs », on comparera par exemple les populations de deux territoires, l'un ayant bénéficié du dispositif et l'autre non, mais disposant par ailleurs de caractéristiques sociodémographiques proches.
- **Évaluation randomisée (ou contrôlée)** : elle s'applique à des dispositifs innovants dont on cherche à mesurer l'efficacité. Elle consiste à définir une population d'individus éligibles et à les répartir de manière aléatoire dans deux groupes : le groupe des bénéficiaires (souvent appelé groupe test) qui aura accès au dispositif et le groupe des non-bénéficiaires (ou groupe témoin). L'intérêt des méthodes randomisées est de permettre une interprétation causale, ce qui n'est pas envisageable pour les autres méthodes. En effet, le tirage aléatoire supprime les biais de sélection et assure la comparabilité initiale des groupes. Les différences observées *a posteriori* sont donc théoriquement imputables au seul dispositif.

Cette approche quantitative remplit le cahier des charges assigné à l'évaluation : déterminer si le dispositif est efficace ou non, avant de le généraliser. Mais elle ne suffit pas quand il s'agit d'expliquer les raisons et mécanismes de la réussite, ou de l'échec. C'est pourquoi nombre de praticiens de l'évaluation forment des équipes pluridisciplinaires pour mettre en œuvre simultanément des approches qualitatives, certes plus classiques, mais complémentaires. Grâce à cette diversité des approches, les évaluateurs sont informés sur les facteurs qui peuvent faire obstacle ou, à l'inverse, favoriser l'efficacité du dispositif. Ces facteurs peuvent être liés au contexte ou encore à la manière dont les individus mettent en œuvre les mesures concernées.

L'utilité de ces investigations qualitatives, la nécessité de planifier en amont les démarches quantitatives, incitent à associer au plus tôt, et de manière forte, les acteurs des dispositifs à la procédure d'évaluation. Ce qui oblige à rompre avec le modèle « pur » de l'évaluateur parfaitement extérieur à son objet.

Des acteurs à mobiliser d'emblée

Les différentes parties en présence gagneraient toutes à intégrer l'évaluation dès la naissance du projet. Du côté des porteurs et des opérateurs, les échanges avec une entité extérieure les conduisent à se poser de nouvelles questions sur leur projet et finalement mieux se l'approprier. Inversement, le fait que ceux-ci soient sensibilisés aux questions de l'évaluation, ainsi que leur adhésion à la méthode proposée, participent à la réussite de la démarche. Dans cette logique, les évaluateurs ne peuvent se borner à appliquer purement et simplement une méthodologie pré-établie, et doivent composer et échanger avec les porteurs de projets tout au long du processus.

Ainsi, un certain nombre d'évaluateurs, au fur et à mesure du déroulement de l'expérimentation, en reviennent progressivement à des démarches proches de la recherche-action ou « évaluation embarquée ». Ces démarches présentent d'autant plus

d'intérêt que l'on a affaire à des expérimentations inédites, et qui présentent un fort degré d'incertitude. Mobiliser une approche participative permet de pointer les dysfonctionnements qui apparaissent au fur et à mesure et de formuler des propositions. Dans cette approche, les différentes formes d'évaluation sont plus complémentaires que concurrentes. Si l'évaluation expérimentale peut être vue comme une exigence incontournable de mesure de l'efficacité d'un dispositif, mais complexe à mettre en œuvre, elle peut être facilitée par l'association avec d'autres démarches.

Le développement de ce type d'approches et celui, concomitant, d'une « culture de l'évaluation » que d'aucuns appellent de leur vœux, supposeraient aujourd'hui de convaincre les décideurs politiques de l'intérêt des évaluations « lourdes ». Un des freins, déjà pointé en 2005 par le rapport de la Cour des comptes, réside dans le décalage entre le rythme de l'évaluation et celui de la décision.

Pour une pédagogie de l'évaluation

L'évaluation est questionnable à tous les niveaux. Des efforts de pédagogie et de communication doivent être accomplis pour assurer sa diffusion.auprès des porteurs de projets et opérateurs, qui, naturellement portés à croire à l'efficacité de leur action, ne voient pas spontanément l'opportunité d'une évaluation. En outre, la méthodologie ne leur est pas toujours familière. auprès des pouvoirs publics et du public ensuite, qui doivent être informés des conclusions des évaluations pour pouvoir en débattre et en tirer les conséquences adaptées.

Évaluation et action politique ne sont pas sur les mêmes temporalités. La première nécessite un recul important quand la seconde exige des réponses rapides. Cependant, à l'heure des rationalisations budgétaires, l'action publique a plus que jamais le devoir de montrer son efficacité et son efficacité, sur des arguments scientifiques. ■

Pour en savoir plus

Evaluation et données longitudinales : quelles relations ? XVII journées d'étude sur les données longitudinales dans l'analyse du marché du travail, Marseille, Relief n°30, mai 2010.

Suivi et évaluation des politiques publiques et des programmes, M. Basle, Economica, 2008.

Les dispositifs d'évaluation des politiques d'aide à l'emploi de l'Etat, rapport public annuel 2004 de la Cour des comptes, février 2005.

À paraître dans la collection Bref, un zoom sur les approches qualitatives en évaluation à partir d'une expérience menée dans le cadre des projets du Fond d'expérimentations pour la jeunesse.

réagir sur
www.cereq.fr

Bref n° 281 ● janvier 2011

Bulletin de recherche
emploi-formation du Céreq

Directeur de la publication

Frédéric Wacheux

Secrétariat de rédaction et mise en page

Elsa Personnaz

**Centre d'études
et de recherches
sur les qualifications**

10, place de la Joliette,
BP 21321,
13567 Marseille cedex 02
T 04 91 13 28 28

www.cereq.fr

Commission paritaire
n° 1063 ADEP.

Reproduction autorisée à condition
expresse de mentionner la source.
Dépôt légal n° 49-459.

Publication gratuite
ISSN 2107-0822

