

CHRONIQUE D'UNE ÉVALUATION RATÉE

Denis Paillard
Université de Toulouse
Toulouse, France

Résumé : L'évaluation présentée rassemble un ensemble de difficultés cumulées en un cas d'école : confusion entre contrôle et évaluation, insuffisance des moyens, ambitions trop fortes, tension entre commanditaires, situation qui a mis à mal tous les principes qui figurent dans la charte de la Société Française de l'Évaluation : pluralité, distanciation, compétence, respect des personnes, transparence, responsabilité, opportunité. Elle rappelle la nécessité de respecter ces principes, côté commanditaire et prestataire, et l'importance des questions d'éthique; elle montre aussi la fragilité des équilibres qui autorisent la mise en œuvre des évaluations; enfin elle interroge l'utilité de disposer d'instances d'arbitrage.

Abstract: The evaluation presented in this article combines a number of the problems outlined in a textbook example: confusion between "control" and "evaluation," insufficient resources, overly ambitious goals, and tension between stakeholders, a situation that led to disregard for all the principles enumerated in the French evaluation society charter: pluralism, distance, competency, respect for individuals, transparency, responsibility, and opportunity. The lessons learned described in this article highlight the need to follow those principles, both on the part of the agency commissioning the evaluation and the evaluator, the importance of ethical issues, and the fragile balance that allows evaluations to be implemented. Finally, the author questions the usefulness of recourse to arbitration proceedings.

■■■■■■ L'évaluation s'est déroulée en 2003 dans une région française et concernait deux programmes de développement régional : un programme européen décliné dans la région dans un document dit DOCUP (DOCUMENT Unique de Programmation), et un contrat pluriannuel d'interventions passé entre l'Etat et la Région, dit Contrat de Plan État-Région (CPER). Le cahier des charges portait sur une évaluation conjointe de ces deux programmes, à mi-parcours,

Correspondance à l'auteur : Denis Paillard, Institut des Sciences Politiques, Université de Toulouse, IEP, 11 rue des Puits Creusés 31000 Toulouse, France; <denis.paillard@sciencespo-toulouse.fr>

l'évaluation intermédiaire du DOCUP étant imposée par l'Europe, et celle du CPER librement décidée par les partenaires.

UN CONTEXTE TROUBLÉ

Un schéma théorique très construit

Ces programmes s'intéressaient tous deux au développement régional, dans toutes ses dimensions (infrastructurelles, économiques, sociales, environnementales, formation, etc.) de façon étroitement imbriquée.

Le Contrat de Plan État-Région 2000–2006 fixait, sur une durée de sept ans, le programme des actions décidées conjointement par deux partenaires principaux (et indépendants l'un de l'autre) : le gouvernement central (État) et un gouvernement régional (la Région), entité administrative qui correspond à une Province, à une Autonomie espagnole ou à un Land allemand.

Le DOCUP permet de mobiliser des fonds européens sur des actions décidées conjointement par les partenaires régionaux. L'initiative des mesures est laissée aux partenaires régionaux, mais la Commission Européenne fixe ses règles et ses priorités. Parmi celles-ci, l'obligation d'aligner, en face des fonds européens, un montant équivalent de fonds publics nationaux (donc fournis par la Région, les Départements, et les différentes collectivités locales), en particulier à travers le CPER.

LE DIFFICILE APPRENTISSAGE DE LA DÉCENTRALISATION

Dans le contexte français, la décentralisation est toujours en construction : les premières lois en la matière datent de 1983, après près de deux siècles d'un centralisme très puissant, caractérisé par une tutelle de l'État central sur des gouvernements locaux atomisés (100 départements et 36 000 municipalités). Les Régions n'ont été créées comme gouvernements autonomes que depuis 1986 (premières élections des conseillers régionaux au suffrage universel).

Dans ce contexte, la construction régionale est encore jeune; cela demande à l'État et ses services de changer de posture et de revoir leurs modes d'intervention; les exécutifs régionaux voient monter en puissance leurs responsabilités, enfin tous doivent faire l'apprentissage d'une contractualisation dans laquelle aucun partenaire ne peut imposer son point de vue, en application du principe de l'indépendance des niveaux de gouvernement.

Cette réalité se traduit dans les documents contractuels et de programmation de la période, qui présentent tous des caractéristiques communes. Ils commencent par des objectifs formulés en termes très généraux et consensuels (réduire de chômage, améliorer la cadre de vie, promouvoir un développement équilibré du territoire, etc.), ce que les évaluateurs ont l'habitude de qualifier d'intentions difficilement évaluables.

Ensuite, ils comportent, rassemblées en axes généralement thématiques, un catalogue de mesures dans lequel certaines correspondent effectivement à des objectifs partagés, et d'autres relèvent d'objectifs propres de chacun des partenaires, juxtaposées dans le document mais conduites indépendamment par chacun.

Les indices permettant de distinguer ces situations sont à la fois financiers (qui est bailleur de fond unique ou principal de la mesure) et organisationnels (qui pilote, qui met en œuvre).

Une conjoncture défavorable, avec un conflit ouvert entre les deux commanditaires

L'évaluation a été engagée à mi-parcours des programmes, dans un contexte particulièrement tendu, à la fois sur le plan institutionnel, politique, et financier.

1. au plan institutionnel, c'était le moment de la discussion et du vote d'une loi de poursuite de la décentralisation, prévoyant un transfert massif de compétences et de personnel vers les collectivités territoriales, à un moment où ces dernières mettaient en avant la disparité entre les transferts de compétences et de charges déjà opérés et les transferts de ressources sensées les compenser, pesant lourdement sur leurs budgets;
2. politique ensuite : la majorité des Régions étaient d'un bord opposé au gouvernement central, et les élections régionales de 2004 approchaient; et
3. financière enfin : les Régions et l'État se reprochaient mutuellement de ne pas avoir respecté leurs engagements financiers dans le Contrat de Plan en cours.

Ces reproches mutuels étaient en outre alimentés par de fortes inquiétudes sur les niveaux d'engagement financier des programmes européens, dont les retards laissaient craindre un « dégageant d'office », correspondant à la perte sèche de l'écart entre les montants

prévus à mi-programme et les montants réels justifiés, chacun rejetant par avance sur l'autre la responsabilité des dégagements.

UNE EVALUATION À REBONDISSEMENTS

La conduite de l'évaluation, qui a duré près d'un an, a été ponctuée de difficultés croissantes, attisées par le manque d'expérience du chargé de mission du commanditaire principal (issu d'un corps d'inspection) qui effectuait là sa première mission sur un programme européen et régional et sa première évaluation.

Le recalage des ambitions

La première étape, fréquente dans les évaluations conduites à l'époque sur les programmes européens, a consisté à réduire les ambitions. La réduction du nombre des questions évaluatives était indispensable, devant près de 50 questions inscrites dans le cahier des charges. Il a donc été admis de procéder à un examen d'ensemble du programme, pour apporter des éléments de réponse générale à chacune de ces questions, puis d'identifier des champs à approfondir, sur lesquels concentrer les efforts.

Plus gênant, il est vite apparu que les systèmes d'information des partenaires étaient hétérogènes et incompatibles : les services de l'État disposaient, sur un outil centralisé, de l'ensemble des données de suivi physico-financier concernant le programme européen (DOCUP), alors que pour le CPER la Région disposait de ses propres données de suivi, et les services de l'État en région avaient chacun (par ministère) leurs propres bases d'opérations. L'organisation des bases de données et les systèmes étaient différents, sans clé commune pour pouvoir rapprocher les opérations. Les seules informations compatibles étaient donc les totaux par mesure, élaborées par les différents services, mais il s'est vite avéré que ces informations avaient un degré de fiabilité assez relatif.

La bataille des méthodes

La première réunion du comité de pilotage s'est bien déroulée; elle portait sur la question des méthodes, de la construction d'un questionnaire d'enquête auprès des porteurs d'enjeu et de bénéficiaires, et ainsi de suite. Le commanditaire principal a cependant exigé de pouvoir réviser le questionnaire d'enquête, tout comme la méthode

d'échantillonnage, mais sa demande de pouvoir contrôler nominativement la composition de l'échantillon a été refusée par l'évaluateur.

Par ailleurs, l'analyse des indicateurs physiques retenus par les partenaires régionaux a rapidement montré que plus du tiers de ces indicateurs n'étaient pas utilisables (non pertinents ou impossibles à renseigner).

La bataille des chiffres

L'examen général des programmes et de leur état d'avancement a rapidement fait apparaître deux difficultés : mesure par mesure, les données agrégées fournies par l'État ne correspondaient pas avec celles de la Région, et qui plus est, dans les deux cas elles ne correspondaient pas non plus au total des opérations figurant dans les fichiers de suivi des programmes, au titre de chacune des mesures, avec parfois des écarts très importants.

Ce constat a conduit l'un des commanditaires à exiger le retrait de l'essentiel des tableaux et analyses du premier document d'étape, la veille de la réunion du comité de pilotage, remettant en cause le travail effectué par les évaluateurs et leur capacité à traiter des données. Pendant la réunion, les deux principaux partenaires ont mis en doute la réalité des chiffres communiqués par l'autre partie.

L'explication, connue plus tard, était la suivante : chacun des services de l'État affectait des opérations à l'un ou l'autre des programmes au gré des exigences de son ministère de tutelle (administration centrale), et les montants totaux étaient ensuite « révisés » en vue des négociations entre État et Région. Les services de la Région procédaient de même.

La bataille des chiffres a connu deux autres terrains d'engagement :

1. l'actualisation des données : un certain nombre d'erreurs furent reprochées à l'évaluateur (ces écarts entre les extractions et la réalité provenaient d'un défaut d'actualisation des données, avec des retards atteignant plusieurs mois dans certains services (inscription des opérations, montants programmés, dépenses, etc.); et
2. l'examen des indicateurs de réalisation et de critères qualitatifs. S'appuyant sur les demandes européennes reprises dans le cahier des charges, les commanditaires exigeaient un traitement et une analyse de ces données. Or, considérés

comme « accessoires », les indicateurs de réalisation physique étaient très faiblement renseignés, avec de nombreuses erreurs. La situation concernant les indicateurs qualitatifs, correspondant à des priorités transversales de l'Europe, était encore moins bonne (e.g., prise en compte de l'égalité des chances femme-homme).

La bataille des analyses : trouver un arbitre

Après six mois de travail, alors que les travaux semblaient avancer à la satisfaction de l'ensemble des parties, et que le consultant en charge du dossier n'avait jamais fait mention de difficultés particulières, la remise du rapport intermédiaire, comportant l'essentiel des constats issus des enquêtes auprès des partenaires et bénéficiaires, et des analyses qui en découlaient, a donné lieu à une mise en cause très violente du travail d'évaluation : contestation de la réalité des entretiens (avec l'exigence de disposer de la liste des personnes rencontrées et des comptes rendus des entretiens), refus des constats et des analyses, mise en doute de la compétence des évaluateurs et, en particulier, du consultant en charge du dossier (en pointant toutes les insuffisances du rapport intermédiaire, par rapport au résultat final attendu, et en notant des approximations et quelques contre-sens).

L'évaluateur a réagi par trois actions :

1. une expertise du dossier, à sa demande, par deux instances extérieures incontestables (spécialistes de l'évaluation de l'administration centrale) permettant de conclure à la conformité des documents aux standards de la profession;
2. le refus de fournir des informations nominatives; et
3. une demande d'arrêter la mission d'évaluation, ou de la poursuivre à la condition que les principes de l'évaluation soient réaffirmés et respectés, tout en remplaçant le consultant en charge du dossier par un autre membre de l'équipe.

Cet épisode s'est conclu par le remplacement du consultant et par un engagement de modération de la part des commanditaires, dans leurs réactions et leurs propos.

Stupeur et tremblements

Un nouveau consultant, d'âge mur, rompu aux situations conflictuelles, a repris l'ensemble du dossier, rectifié des erreurs matérielles,

et poursuivi les entretiens, le recueil, et l'exploitation des données au cours de l'été. En septembre, une première version du document final a été envoyée aux membres du comité de pilotage. Alors que le consultant s'était fortement investi sur le dossier, il a brusquement demandé à en être déchargé.

Une investigation rapide a alors révélé des pressions de la part d'un des commanditaires, pour orienter les conclusions, assorties de menaces sur un possible licenciement du consultant « comme le précédent » en cas de refus (lequel avait effectivement été licencié, mais pas sous la pression du commanditaire).

Apothéose

La remise du rapport final a donné lieu à une ultime difficulté : envoyé par courriel pour relecture aux deux commanditaires principaux, le rapport est revenu avec une exigence de modifier certaines conclusions et d'en supprimer d'autres et, de la même manière, de modifier certaines recommandations ou d'en supprimer. Plusieurs chapitres avaient été amputés ou réécrits, notamment le compte rendu d'exécution de la mission.

Après réception d'un message de l'évaluateur indiquant que le commanditaire n'avait pas le pouvoir d'infléchir la nature des conclusions et des recommandations, accompagné d'un fichier verrouillé, l'incident s'est clôt au cours d'une réunion à niveau exécutif (État + Région) au cours de laquelle les commanditaires ont évoqué les craintes de faire « mauvaise figure » au niveau national et devant les instances européennes, des risques potentiels sur les financements futurs, ainsi que ceux d'une mauvaise utilisation du document au sein de la région, dans le cadre de la campagne électorale.

Finalement quelques reformulations ont été effectuées sans modification des contenus.

ENSEIGNEMENTS

L'application des principes de l'évaluation

Si l'on reprend les principes de la Charte de l'évaluation adoptée par la Société Française de l'Évaluation (2006), on peut noter qu'aucun des principes figurant dans cette charte n'a été respecté :

- la *pluralité*, le commanditaire exigeant de définir lui-même l'échantillon, puis refusant des points de vue distincts du sien, dont ceux des bénéficiaires, y compris celui de l'évaluateur;
- la *distanciation*, avec la confusion entre les fonctions de gestion des programmes, la négociation entre les deux commanditaires, et le pilotage de l'évaluation; enfin et plus grave, les interventions et les pressions pour orienter les conclusions;
- la *compétence*, principe écorné à la fois par le commanditaire, nommant un chargé de mission ne connaissant pas le programme et n'ayant jamais pratiqué d'évaluation, et par l'évaluateur (son premier consultant);
- le *respect des personnes*, avec une demande de levée de l'anonymat des personnes enquêtées et de leurs propos, et de respect des évaluateurs par les commanditaires ;
- la *responsabilité*, avec une confusion des rôles des commanditaires, une absence de mandat clair, l'inexistence de systèmes d'information adéquats conduisant à une incapacité à fournir des informations de base fiables et exploitables;
- l'*opportunité*, enfin qui n'a pas été validée : la situation de conflit ouvert entre les deux principaux commanditaires sur l'un des deux programmes (Contrat de Plan) aurait du conduire à ne proposer qu'une évaluation du programme européen (DOCUP). Il faut noter que ce principe d'opportunité a été introduit plus tard (2006) dans la charte de la SFÉ.

Les raisons habituelles d'un échec prévisible

Les difficultés rencontrées dans cette évaluation sont finalement, en partie, banales. La démesure des ambitions est flagrante, tout comme leur inadaptation, à la fois à l'état des systèmes d'information (existence de systèmes d'information, formats et possibilités d'extraction, fiabilité des données, et taux de renseignements) et aux moyens consacrés à l'évaluation. La difficulté à hiérarchiser les questions et la juxtaposition des demandes des partenaires y joue également un rôle.

Une expertise préalable des systèmes d'information, l'annonce claire dans la proposition du nombre de questions auxquelles il est possible de répondre, en fonction des budgets et des méthodes, enfin la systématisation d'une phase de (re)cadrage, au démarrage des travaux, sont de nature à résoudre ces difficultés.

Une incompréhension de l'exercice d'évaluation est la seconde source de difficultés. La confusion avec les activités de contrôle ou d'audit, l'absence de culture de l'évaluation, et la crainte de la sanction militent pour donner une place centrale à la dimension d'apprentissage de l'évaluation, quitte à en restreindre les ambitions.

Les questions d'éthique au centre des difficultés

Se pose clairement, à travers cet exemple, une question d'éthique. Sur un plan général, la situation est un bon exemple de la fâcheuse habitude de certains commanditaires (publics ou privés) à vouloir infléchir les résultats des travaux des consultants qu'ils mandatent pour des travaux supposés indépendants, et de la difficulté des prestataires à y résister (et parfois de leur complaisance).

Dans l'évaluation, la nécessité de la rigueur éthique est d'autant plus importante que les tentations sont accentuées par les enjeux et les craintes liées à l'exercice. Une référence explicite aux principes de l'évaluation (e.g., la charte) dans les appels d'offre et dans les propositions est de nature à prévenir ces risques. Un « mode d'emploi » de ces chartes pourrait être utile, en détaillant comment doivent être appliqués les dits principes, et en explicitant les règles éthiques.

L'exemple décrit ici pose la question du respect des principes et des règles par les parties dans le déroulement des travaux, et l'utilité d'une instance d'arbitrage en cas de conflit majeur. *A contrario*, il indique les limites de l'exercice dans des situations où l'un des partenaires exerce un pouvoir sur les autres (systèmes hiérarchiques, situations non démocratiques).

La situation de conflit et la question de l'opportunité

L'expérience décrite interroge clairement le principe d'opportunité, qui ne figurait pas dans la charte de la SFÉ au moment des faits.

Est-il possible de réaliser une évaluation conjointe pour deux partenaires en conflit majeur? La réponse est probablement négative, sauf à faire acter, dès le début de l'exercice, des règles de respect de la neutralité de l'évaluateur dans ce conflit. Dans ce cas, l'évaluation pourra servir à (re)construire des bases et référentiels communs permettant d'assoir une négociation qui doit se développer dans un autre cadre.

CONCLUSIONS : UNE ÉVALUATION RATÉE (DANS SA CONDUITE) MAIS UTILE (DANS SES RÉSULTATS)

L'évaluation a fourni une image non déformée du programme et de son déroulement, et des recommandations à la fois stratégiques et opérationnelles, au prix de nombreuses difficultés et de conflits parfois violents.

La réalisation de l'évaluation a par ailleurs mis en évidence des dysfonctionnements des dispositifs de suivi et de pilotage des programmes, conduisant à la mise en place rapide de mesures correctives, sur la structure de ces dispositifs, la mise à jour des données, et leur utilisation pour le pilotage opérationnel.

La présentation finale a rassemblé un comité de pilotage élargi, avec la plupart des services techniques tant de l'État que de la Région, ainsi que des partenaires extérieurs (Départements), lequel a discuté au fond les conclusions et les propositions. Les gestionnaires techniques des mesures y ont souligné l'utilité de disposer d'un référentiel commun, l'apport de la pluralité des points de vue, et de conclusions a priori dérangeantes. Par-delà les résultats immédiats, cette évaluation a finalement apporté sa contribution à l'apprentissage collectif et à la diffusion de la culture de l'évaluation.

RÉFÉRENCES

Société Française de l'Évaluation. (2006). Charte de l'évaluation des politiques publiques et des programmes publics. Repéré de http://www.sfe-asso.fr/sfe-evaluation.php?menu_id=191

Denis Paillard, ingénieur agronome et économiste, poursuit des activités à la fois comme consultant sur les politiques de développement durable (rurales, territoriales, sectorielles) et leur évaluation, et comme enseignant en stratégies de développement, ingénierie de projets, et évaluation des politiques publiques; il collabore régulièrement aux activités de la Société Française de l'Évaluation dont il a été vice-président, et au Réseau francophone d'évaluation.