

La référentialisation : une façon de modéliser l'évaluation de programme, entre théorie et pratique

Vers une comparaison des approches au Québec et en France

Gérard Figari

Claire Tourmen

Université Pierre-Mendès-France, Grenoble

MOTS CLÉS: Activité évaluative, programme, «référentialisation»

Nous proposons de distinguer, en nous inspirant des théorisations de l'évaluation, des éléments invariants qui seraient constitutifs de l'activité évaluative. Nous tentons ensuite de comparer la façon dont ils sont utilisés et nommés dans des pratiques d'évaluation de «programme/dispositif», en France et au Québec. Nous proposons enfin une modélisation de l'évaluation de «programme/dispositif», qui touche à la fois à l'objet et à l'activité d'évaluation sur la base de ses invariants (démarche théorique) et qui peut se décliner, dans la pratique, sous forme de protocole méthodologique : la «référentialisation».

KEY WORDS: Evaluative practices, program, «référentialisation»

Our purpose is to identify invariant elements describing evaluative practices within the evaluation theories. We try then to compare the way they are used and named in «program» evaluation practices in France and Quebec. We finally propose a model, called «référentialisation», of evaluation's object and practice based on these invariant elements. It can also be used as a method to evaluate.

PALAVRAS-CHAVE: Atividade avaliativa, programa, «referencialização»

Propomo-nos distinguir, inspirando-nos nas teorizações da avaliação, os elementos que serão constitutivos da atividade avaliativa. Seguidamente, tentamos comparar a forma como eles são utilizados e nomeados nas práticas de avaliação de «programa/dispositivo», em França e no Québec. Finalmente, propomos uma modelização da avaliação de «programa/dispositivo», que toca simultaneamente o objecto e a actividade de avaliação, na base das suas invariantes (procedimento teórico) e que pode declinar-se, na prática, sob a forma de protocolo metodológico: a «referencialização».

Introduction

Évaluer des «programmes», des «dispositifs», des «politiques publiques», des «projets»... Si nous postulons que ces termes différents recouvrent des «objets» très proches que l'on désignerait, selon les cultures, par la notion de «programme» en Amérique du Nord et par la notion de «dispositif», notamment en France, nous disposons alors d'une remarquable occasion de comparaison. C'est sur ce postulat d'existence d'un objet, qu'on appellera provisoirement le «programme-dispositif», que nous allons conduire notre réflexion. Dans ce cadre, la diversité des approches et des concepts utilisés, au lieu de constituer le signe d'une confusion sémantique, peut devenir une source de confrontation théorique, méthodologique et pratique: tel est l'enjeu de cette contribution.

La comparaison de pratiques d'évaluation de «programmes-dispositifs» demande, tout d'abord, une clarification des concepts sous-jacents. Qu'est-ce qu'évaluer un «programme-dispositif»? Nous ferons l'hypothèse que, comme toute forme d'évaluation, celle-ci se déroule selon des invariants qui se retrouvent quel que soit le type d'évaluation menée et quel que soit l'objet: il s'agit donc de commencer par un repérage de ces invariants, qui seraient des dimensions toujours présentes dans les activités évaluatives.

En second lieu, nous allons esquisser une réflexion comparative entre des approches québécoises et françaises. Il s'agit de tenter une comparaison entre les conceptions du rapport «théorie-pratique» en matière d'évaluation de «programme-dispositif» dans les deux contextes, avec l'hypothèse d'un possible enrichissement mutuel. Nous allons notamment comparer la façon dont les invariants de l'activité évaluative sont nommés et pratiqués dans ces deux contextes, sur la base de lectures comparatives et d'entretiens avec des évaluateurs.

Enfin, nous essayerons d'appliquer cette explicitation des éléments communs repérés à la présentation d'une modélisation de l'évaluation de «programme-dispositif», la «référentialisation» (Figari, 1994), qui revêt un double statut: elle modélise à la fois l'objet et l'activité d'évaluation sur la base de ses invariants (démarche théorique) et peut se décliner, dans la pratique, sous forme de protocole méthodologique. Cette réflexion pourrait constituer une contribution à des projets de recherche communs.

À la recherche des invariants de l'activité évaluative

Comment comparer les différentes activités évaluatives ?

Tout d'abord, la documentation sur l'évaluation nous permet de poser une série de postulats concernant ce que nous appelons des « invariants » de l'activité évaluative. Certains auteurs ont en effet tenté de décrire la façon dont les activités se déroulent, et cette forme de modélisation nous autorise à supposer que l'activité évaluative se déroulerait selon certains invariants, à savoir des opérations toujours présentes quelle que soit leur forme concrète de réalisation et d'explicitation.

Pourtant, dégager un modèle du déroulement de l'activité évaluative est à première vue difficile. Le terme « évaluation » semble être une généralisation qui masquerait la diversité des pratiques. En effet, de nombreuses activités ont pris, au fil du temps, l'appellation « évaluation », dans différents champs : on évalue les apprentissages des élèves dans la classe, les acquis d'expérience (RVAE), les compétences au travail, la réussite d'un dispositif de formation, les effets d'un « programme », d'un « dispositif » ou d'une « politique publique », etc. De plus, les ouvrages sont produits dans des champs disciplinaires différents. Peut-on alors dire qu'évaluer est une activité spécifique ? Existe-t-il une unicité de ces pratiques au-delà des différents contextes et des différents objets évalués ?

Dans la documentation sur l'évaluation, on trouve de multiples définitions de ce que seraient des pratiques d'évaluation. Sont-elles généralisables à toute activité d'évaluation ? Et laquelle retenir, au risque de faire un choix restrictif ou normatif ? Nous avons choisi de suivre certains auteurs qui ont tenté une telle généralisation de la définition de l'évaluation pour en extraire les caractéristiques essentielles. Ainsi, Stufflebeam, Nadeau, Hadji, Figari, Barbier ont avancé dans cette voie.

En tentant de décrire l'activité évaluative, ils ont mené une comparaison entre ses différentes formes, y compris la plus « spontanée », pour reprendre un terme de Barbier (1985). Il s'agit de celle que nous menons tous dans notre vie quotidienne. Stufflebeam (1980) précise que, dans tout jugement évaluatif, on retrouve la même structure, même si celle-ci reste implicite comme dans les jugements rapides de la vie quotidienne. « Parce qu'il est implicite, ce processus garde secrets à la fois les données considérées et les critères ou standards qui ont servi à les apprécier » (*id.*, p. 17). Nadeau (1988) dit aussi à propos des jugements courants et informels : « Ces évaluations journalières

sont informelles, ce qui ne veut pas dire qu'elles soient inexactes, en ce sens qu'elles sont le plus souvent basées sur des impressions vagues et diffuses et sur des normes intuitives.» «À moins que ces évaluations ne soient faites à partir de critères bien définis et sur la base d'informations pertinentes et valides, elles ne rencontrent pas le véritable sens du concept "évaluation formelle" (*id.*, p. 35). Tout ceci laisse à penser que l'activité évaluative posséderait quelques éléments invariants qui seraient présents même quand elle est pratiquée de façon non professionnelle, quotidienne, ce qui peut éclairer les constituants de l'évaluation «formelle» (Nadeau).

Les invariants de l'activité évaluative

Nous retiendrons tout d'abord que l'activité évaluative contribuerait à produire des données permettant l'attribution de jugements, finalisés par des visées sociales externes au processus comme la prise de décision. Stufflebeam (1980), après avoir comparé différentes définitions de l'évaluation, comme la mesure, la congruence des résultats avec les objectifs (Tyler, 1950) et le jugement, conclut que la seule collecte des données ne suffirait pas pour définir l'évaluation mais qu'il y aurait toujours préparation d'un jugement de valeur, qui poserait par exemple la question : *dans quelle mesure* les objectifs ont-ils été atteints? Hadji (1989) affirme, quant à lui, que le produit de l'évaluation est toujours exprimé sous forme de jugement. En logique, celui-ci pourrait être défini comme une «assertion qui établit ou dément un rapport entre deux termes, pose un rapport comme vrai ou faux» (Hadji, 1989, p. 28). Il peut s'agir d'un jugement de valeur ou de réalité (qui énonce un fait puis qui apprécie ce que vaut la réalité). Les deux auteurs insistent sur le fait que ces jugements sont finalisés par des visées externes au processus qui le justifient, comme la prise de décision.

Nous retiendrons aussi que, dans l'activité évaluative, les jugements seraient produits en comparant des données («référés») à des «réfèrents», permettant d'affecter des attributions causales expliquant les phénomènes observés. Deux conceptions distinctes de la notion de réfèrent sont à considérer.

- Celle de Hadji qui la formule ainsi : «Nous proposons de nommer évaluation l'acte par lequel on formule un jugement de "valeur" portant sur un objet déterminé, par le moyen d'une confrontation entre deux séries de données, qui sont mises en rapport : des données qui sont de l'ordre du fait [...], des données qui sont de l'ordre de l'idéal (attentes, intentions, projet sur le même objet)» (Hadji, 1989, p. 25). Il ajoute : «L'essentiel de l'évaluation réside dans une mise en rapport entre ce qui existe et ce qui était

attendu, entre une réalité et un idéal.» Ces descriptions du processus amènent à distinguer les «référés», à savoir les données sur l'état de l'objet évalué, et les «référents», système de référence permettant de lui attribuer un sens.

- Celle de Figari qui s'est opposé à une conception trop restrictive de la notion de «référent» se limitant aux seules normes ou idéaux à atteindre. Il rappelle que, comme en linguistique (où la notion de référent désigne «un élément extérieur» par rapport auquel un objet est évoqué) ou comme en mathématiques (où le référentiel cartésien consiste à définir la situation d'un point par rapport à des axes de coordonnées), le «référent» de l'évaluation est une «information choisie en référence». Il peut être constitué par un idéal ou une norme mais aussi par un niveau, un élément de comparaison particulier, non normatif. Il parle alors de «système de références», au pluriel, soulignant qu'«il n'est pas envisagé d'évaluation sans recours implicite ou explicite à un système de références» (Figari, 1994, p. 31).

Nous retiendrons enfin que, dans l'activité évaluative, un processus de (re)construction des référents et des «référés» serait à l'œuvre pour permettre de formuler des interprétations. Dans l'évaluation, les «référés» sont la plupart du temps produits. Une opération de production de données sur l'état de l'objet évalué est en effet souvent à l'œuvre. Elle vise à dépasser une connaissance partielle ou insuffisante sur l'objet évalué, ce qui a pu amener à assimiler l'évaluation à une seule activité de production de connaissances.

Évaluer consisterait ainsi à se représenter son objet en le «modélisant» (Figari, 1994). On retrouve exprimée l'idée que les objets évalués ne sont pas donnés mais construits par les évaluateurs. Selon Figari (1994), l'évaluateur modélise l'objet évalué, ce modèle lui permettant de maîtriser sa représentation de l'objet afin de s'orienter face au foisonnement des phénomènes réels. Différents outils sont ainsi proposés aux évaluateurs pour modéliser un «programme/dispositif», un «projet» ou une «politique publique» (CIPP de Stufflebeam, ICP de Figari [1994], mais aussi le modèle logique...) et recueillir des données à son égard.

Si les référés sont produits au cours de l'activité, qu'en est-il des référents? Hadji précise que les référents sont, en quelque sorte, préalables à l'activité évaluative: il parle d'un «modèle idéal qui oriente la lecture de la réalité et qui préside à la prise d'indices» (Hadji, 1989, p. 42). Figari explique que la référence est une action consistant à désigner des repères préexistants destinés à situer l'objet à évaluer, par rapport à ces repères (Figari, 1994, p. 13).

Or, il semble que l'activité évaluative se caractériserait par la construction – ou la reconstruction – des référents. Comme le dit Barbier, l'activité d'évaluation consiste à produire des «référés» mais aussi des «référents» (Barbier, 1985, p. 83). Il note que, dans l'évaluation, certains objets deviennent des «référents» et qu'«un objet ne devient référent qu'au sein de l'acte d'évaluation» (p. 73). «Comme les données de l'évaluation, les critères ou les normes de jugement n'existent pas à l'état naturel; ils sont toujours le résultat de processus spécifiques conduisant à leur élaboration.» Ces processus ne seraient explicités que lors des évaluations «instituées». C'est ce qu'il appelle «la tâche d'élucidation des référents» et de «spécification des critères» (p. 77). Il observe qu'elle se matérialise souvent sous la forme d'explicitation des objectifs du dispositif évalué, qui sont ses référents classiques.

L'activité de production des référents consisterait ainsi en la spécification des éléments de référence qui vont guider l'interprétation des données sur l'objet (ce sont des objectifs, des valeurs, d'autres données, des textes d'orientation ou même des «référentiels» officiels, etc.).

Sous réserve de vérification, nous faisons l'hypothèse que, en ce qui concerne ces invariants issus des théorisations de l'évaluation, la similitude des catégories de questions soulevées par l'évaluation, au Québec et en France (et plus généralement, en certains cas, en Europe francophone) est établie.

L'évaluation de programme au Québec et en France

La modélisation de ces invariants permettra-t-elle de comparer les approches d'évaluation de programme et de dispositif en France et au Québec? Nous postulons qu'elles traitent chacune ces invariants de façon spécifique.

Des préconisations comparables

Tout d'abord, les pratiques sont nommées de différentes façons en France et au Québec. D'un côté de l'Atlantique, en France et dans certains pays d'Europe, il est coutume de parler des pratiques «d'évaluation de politiques publiques», les objets évalués pouvant être des «dispositifs», des «politiques publiques», des «projets» ou des «programmes». Nous choisirons comme exemple, ici, «l'évaluation des politiques publiques» comme un des éléments comparables à un type de programme.

Le guide du Conseil scientifique de l'évaluation (CSE, 1996), un des premiers en la matière et considéré comme une référence en France, définit l'évaluation de politiques publiques comme des «évaluations de grande

ampleur, décidées par le gouvernement concernant des politiques complexes à objectifs et acteurs multiples», avant d'ajouter que l'objet de cette activité peut être plus vaste : «Le présent guide est consacré à l'évaluation des "actions" et "activités" publiques, vocables délibérément imprécis pouvant désigner une large gamme d'objets : politiques, programmes, actions, mesures, réglementations, projets, activité d'une administration et d'un corps de fonctionnaires, services produits par un organisme ou un service public, etc. L'évaluation individuelle des personnels est en revanche ici clairement exclue.» Le terme «politique» est ainsi employé «par commodité», c'est une «référence habituelle dans le contexte français» qui a un «sens générique» et peut inclure les différents objets cités. Cette forme d'évaluation concerne donc des objets vastes : le terme de «politique» renvoyant à l'existence d'un projet incluant de multiples objectifs, de multiples acteurs et sans temps déterminé *a priori*; elle concerne aussi des objets plus ciblés comme des «programmes», cette notion faisant référence aux pratiques de programmation budgétaire, un programme se définissant par des objectifs ciblés, des moyens ciblés et un calendrier précis, (*cf.* les pratiques de la Commission européenne).

Au Québec, par contre, il est plus courant de parler «d'évaluation de programme» même si cette forme d'évaluation peut aussi concerner des «politiques» (au sens de «policy») ou des «projets» plus restreints. Le document d'orientation du Secrétariat du Conseil du trésor (2002) définit l'évaluation de programme comme suit : «L'évaluation de programme est une démarche rigoureuse de collecte et d'analyse d'information qui vise à porter un jugement sur un programme, une politique, un processus ou un projet pour aider à la prise de décision.» La notion de «programme» est par ailleurs définie comme telle : «Le programme désigne ici un ensemble cohérent et structuré d'objectifs, d'activités et de ressources humaines, financières, matérielles et informationnelles, regroupés pour offrir des produits et des services particuliers qui satisfont les besoins ciblés d'une population prédéterminée.» Hurteau et Houle, inspirés par les travaux de Chen, donnent une définition encore plus extensive au terme de programme : «Un programme constitue ainsi un système ouvert et systématique composé des dimensions qui lui sont propres – intrants, processus et résultats, environnements et rétroactions.» (Hurteau & Houle, 2005, p. 86.) On voit donc que le terme «évaluation de programme» regroupe l'ensemble des objets de cette forme d'évaluation, même si le terme «programme» peut être restreint à un sens gestionnaire proche de celui utilisé en France et dans la Communauté européenne.

Au-delà de ces différentes catégorisations des objets évalués, existe-t-il des préconisations comparables? Dans les deux cas, ce type d'évaluation est défini par ses finalités, qui sont de concourir à l'amélioration des pratiques de gestion publique par la production de connaissances sur la mise en œuvre et les effets des interventions publiques, le processus étant finalisé par l'attribution de valeur, ce qui le différencie d'autres types d'études et de recherches. Ainsi, le guide du CSE définit le but de l'évaluation comme suit: «[...] fondamentalement, l'évaluation consiste à tenter de répondre à un ensemble de questions relatives à une politique». Il souligne qu'elle peut servir différentes finalités, comme la finalité «gestionnaire», «décisionnelle», «démocratique», etc. Le document d'orientation du Secrétariat du Conseil du trésor du Québec (2002) souligne, quant à lui, que «l'évaluation de programme sert principalement de soutien au processus de décision et de planification et vise l'optimisation des ressources». Dans les deux cas, l'évaluation de «programme-dispositif»¹ est bien un processus de production de connaissances et d'attribution de valeur servant différentes finalités externes, liées à des préoccupations de gestion publique.

Comment sont produits ces bénéfices attendus de l'évaluation (aussi appelés «extrants» au Québec)? Dans les deux cas, il semble que le processus est orienté par des questions d'évaluation auxquelles il va chercher à répondre. En effet, différents guides de référence français et européens (CSE, 1996; MEANS, 1999) soulignent qu'une des premières étapes consiste à préciser les finalités de chaque évaluation et à formuler des questions sur l'objet évalué. Il est précisé que ces questions peuvent porter sur différents critères, les plus cités étant «cohérence, efficacité, efficience, impact, pertinence...» (CSE, 1996). De son côté, le document d'orientation du Secrétariat du Conseil du trésor du Québec (2002) précise que «l'évaluation de programme peut traiter les champs suivants: la pertinence, l'efficacité, l'efficience, l'impact» et porter sur les «besoins, les possibilités d'action, l'implantation ou les résultats et les effets». Ainsi, dans les deux cas, les préconisations méthodologiques concernent la définition de questions d'évaluation qui mettent l'accent sur l'un ou l'autre des critères cités, comme autant de façons d'examiner les objets évalués.

Les textes prescriptifs des deux pays insistent de plus sur le fait que les «objectifs» des «programmes-dispositifs» évalués doivent être déterminés au préalable afin de préparer les opérations de collecte et de traitement des données. Les objectifs acquièrent ainsi le statut de référents légitimes, les plus utilisés pour évaluer. Toutefois, le document d'orientation du Secrétariat du Conseil du trésor du Québec (2002) précise que l'évaluation peut être «axée

sur les résultats» (en lien avec les objectifs) mais aussi sur «la conformité avec le besoin exprimé» par les citoyens, ce qui laisse la porte ouverte à l'utilisation d'autres référents (les «besoins») même si l'accent semble davantage porté sur l'axe résultats/effets. De son côté, le guide du CSE (1996) insiste sur l'élaboration du «référentiel» d'évaluation qui serait même «l'une des questions majeures» à l'étape du projet, permettant de déterminer des critères et des normes de référence. Il souligne qu'à première vue, les objectifs de la politique évaluée sont le seul référentiel légitime, en référence au décret du 22 janvier 1990 instituant l'évaluation de politiques publiques en France; toutefois, les auteurs observent que des problèmes peuvent surgir en pratique: il faudrait alors prendre en compte les objectifs implicites, tous les objectifs n'étant ni de même niveau ni cohérents, leur formulation étant évolutive et tous n'étant pas directement observables. Il faudrait donc «interpréter les objectifs de la politique en fonction du contexte et des priorités actuelles du commanditaire», ce qui est nommé phase de «contextualisation». On voit que, dans les deux cas, les guides de référence rappellent la nécessité de déterminer – ou de reconstruire – les «référents» servant à l'évaluation (objectifs politiques, «besoins» ou éléments contextuels) même si le terme de «référent» est lui-même peu employé.

Les pratiques : variations et similitudes

Après avoir fait un bref tour d'horizon des documents de référence en évaluation de «programme/dispositif» en France et au Québec, nous allons examiner les pratiques telles qu'elles nous ont été décrites. Nous avons mené des entretiens avec des professionnels exerçant des activités d'évaluation de politiques publiques ou de programme à temps plein, sur les deux terrains, dans différents postes de différentes Administrations ou de cabinets privés. Nous avons ainsi rencontré sept professionnels au Québec et 16 en France et à la Commission européenne. Ces entretiens ont été menés selon une grille qui visait à susciter une description des postes occupés, des parcours et de l'activité de travail habituellement réalisée ainsi que de ses difficultés.

Une dénomination variable des professions

Un des premiers constats est que les postes de ces professionnels n'ont pas de dénomination fixe. Les activités d'évaluation peuvent être mêlées à des activités de «contrôle de gestion» (en France et à la Commission européenne) ou de «vérification interne» (au Québec), de suivi, de planification ou de prospective, etc. Ainsi, d'un côté, beaucoup ne donnent pas de nom à leur poste, certains disent même l'ignorer. Il y a toutefois quelques dénominations

qui reviennent comme «chargé d'étude évaluation», «chargé de mission évaluation» ou «responsable de l'évaluation». En France, par exemple, les postes se répartissent surtout entre des chargés d'évaluation dans l'Administration, en charge de la formalisation de la commande et du suivi des évaluations, plus rarement de leur réalisation (évaluateurs internes), et des évaluateurs qui réalisent ces évaluations, le plus souvent appartenant à des laboratoires de recherche ou des cabinets de conseil privés. D'un autre côté, au Québec, les évaluations semblent plus fréquemment réalisées au sein des ministères par des équipes d'évaluateurs qui sont pilotés par des «coordinateurs d'évaluation» aussi appelés «gestionnaires», même s'il peut être fait appel à des évaluateurs externes.

Une première étape commune : l'orientation de l'évaluation

Au-delà de ces différences d'organisation des fonctions d'évaluation, on observe que les professionnels décrivent tous une première étape d'orientation du travail qui consiste à préciser les finalités spécifiques à chaque évaluation, en lien avec la commande, puis à construire des questions d'évaluation spécifiques qui orientent la suite du travail (collecte et analyse des données). Elle peut être faite au sein de l'Administration commanditaire et discutée avec les évaluateurs. Cette étape est aussi celle où des critères d'évaluation sont définis, ainsi que, la plupart du temps, des indicateurs qui vont guider la collecte des données. Enfin, c'est aussi souvent lors de cette étape que les objectifs de la politique ou du programme évalué ainsi que ses réalisations sont décrits, voire reconstruits avec des outils comme le diagramme logique des impacts ou le modèle CIPP de Stufflebeam (1980). On observe donc que les professionnels, en réalisant différentes productions préalables comme le projet d'évaluation, le cahier des charges ou le mandat, le devis, etc., sont bien dans une démarche d'évaluation où ils préparent l'attribution d'un jugement en précisant quelles questions, quels critères et quels référents vont guider la démarche. Vient ensuite une étape de collecte et d'analyse des données qui a lieu dans toutes les pratiques décrites, menant à la production d'un rapport d'évaluation.

Les difficultés : rigidité des méthodes

Toutefois, ces démarches ne sont pas sans difficultés. La première, qui a été soulignée par la plupart des interviewés, consiste à adapter ces méthodes à chaque contexte d'évaluation, à savoir aux caractéristiques de l'objet évalué, aux attentes et finalités spécifiques et aux ressources disponibles. Ainsi, une des limites de certaines pratiques serait de «plaquer» des méthodes toutes faites pour produire l'évaluation, sans avoir distingué les contraintes et les

ressources propres de chaque contexte. Ce «pilotage par les méthodes» a déjà été évoqué par Chen (1990), qui observait que trop d'évaluateurs sont guidés par des considérations de méthode, en fonction de leur savoir-faire technique, ce qui les conduit par exemple à utiliser toujours les mêmes critères quels que soient les objets et les demandes. Chen préconise de réaliser des évaluations «theory driven», qui partiraient d'un effort préalable de conceptualisation de l'objet à évaluer et de formalisation des référents à partir duquel il est évalué (*cf.* le statut théorique et méthodologique de la «référentialisation», dans Figari, 1994, p. 54, *infra*). Dans ce sens, plusieurs professionnels québécois ont jugé trop «rigide» le document d'orientation du Secrétariat du Conseil du trésor (2002) qui restreint l'évaluation à l'utilisation de quatre critères alors que d'autres critères peuvent être utilisés, de même qu'en France, plusieurs professionnels ont fustigé l'utilisation trop systématique du critère d'efficience et d'indicateurs préformatés, parfois préconisés par la Commission européenne, en raison d'un manque de souplesse pour répondre à chaque demande d'évaluation.

Les difficultés : jugements peu fondés

Une autre difficulté réside dans le fait de pouvoir porter un jugement au terme de l'analyse des données: Hurteau et Duval (2005) ont estimé que c'était là une des limites courantes des rapports d'évaluation de programme. En effet, selon un échantillon d'évaluations de programme qu'elles ont étudié, dans plus d'un tiers des cas, les standards d'évaluation (ou référents et leurs normes de réussite) ne sont pas explicités et l'on ne sait pas au nom de quoi le jugement est porté. Si les éléments sont cités dans un peu moins de deux tiers des cas, peu d'études spécifient ceux qui ont permis de mener la comparaison. Ce serait là une des difficultés réelles de l'évaluation de «programme/dispositif»: réussir à fonder et justifier les jugements portés au regard de la démarche méthodologique. Cette difficulté a été évoquée par les professionnels rencontrés, qui soulignent la difficile question de l'attribution des effets observés à la politique évaluée au regard des nombreux facteurs contextuels potentiels. À partir de quand un programme peut-il être jugé *suffisamment* satisfaisant? Cette question du seuil est redoutable pour les professionnels qui peuvent utiliser des techniques pour tenter de la traiter (grilles multicritères et grilles de notation et de pondération, benchmarking, panel d'experts, etc.).

Ce bref détour à travers les pratiques en France et au Québec ainsi que les textes qui les encadrent permet déjà de mettre en avant une idée importante qui traverse les deux cultures de l'évaluation de programme : celle de la liaison entre théorie et méthode dans la conduite de l'évaluation de «programme-dispositif». Cette idée va sous-tendre l'ensemble de la démarche qui va suivre.

Une modélisation de l'évaluation de «programme-dispositif», la référentialisation

Jusqu'ici, nous avons dressé l'inventaire des invariants de l'activité d'évaluation de «programme-dispositif» et tenté de montrer la comparabilité des concepts, notions et pratiques qui la décrivent. Quant aux instruments proprement dits, ils sont, dans tous les cas, reconnaissables du point de vue de la méthodologie des sciences sociales. Mais si nous en restions à ce simple constat, nous n'aurions pas avancé d'un pas car on sait bien que les pratiques d'évaluation peuvent utiliser les mêmes concepts et notions et cependant se montrer différentes et hermétiques les unes aux autres. C'est pourquoi il nous a semblé utile de tenter d'intégrer les réflexions qui précèdent à un protocole qui utilise tout ou partie de l'appareil théorique inventorié au moyen d'une modélisation de l'évaluation de «programme-dispositif», qui va être exposée ici à travers la «référentialisation» (Figari, 1994). Nous allons partir d'un ancrage théorique de cette forme de modélisation, puis situer la fonction de ses différents descripteurs que sont les référents, critères et indicateurs, pour définir plus précisément la référentialisation avant de terminer par sa visualisation synthétique et un exemple résumé de son application.

Justification théorique d'une modélisation

La documentation portant sur l'évaluation de programme (Shadish, Cook & Leviton, 1991; Younès, 2006) laisse entrevoir des conceptions, théorisations et méthodologies fort différentes qui attribuent aux résultats de cette forme d'évaluation des significations, des productions de connaissances également fort différentes.

Ainsi, selon Younès, les chercheurs ne privilégient pas tous le même type de connaissance à produire : par exemple, «la tradition objectiviste représentée par Campbell et Scriven privilégie la connaissance causale. À l'opposé, un paradigme initié par Stake (1980) puis développé par Guba et Lincoln (1981), Patton (1980), Cranton (2001) qui s'enracine dans une épistémologie constructiviste, interprétative et herméneutique, privilégie la connaissance utile, pluraliste, la découverte. [...] Cronbach et Rossi visent, par l'évaluation, la

construction de connaissance et l'utilisation de la connaissance». Rossi et Freeman (1985, p. 38) définissent comme «compréhensives» des évaluations qui réalisent, à la fois, une «conceptualisation» et un «modèle d'intervention», un «contrôle de l'implantation du programme» et «l'évaluation de l'utilité du programme». L'évaluation compréhensive est «l'application systématique des procédures de recherche en sciences sociales en estimant la conceptualisation et le modèle, l'implantation et l'utilité des programmes d'intervention sociale» (Rossi & Freeman, 1985, p. 19). Younès relève que, pour ces auteurs, la recherche en évaluation ne se limite pas à l'application de méthodes d'appréciation de phénomènes ou de résultats: «C'est aussi une activité politique et managériale, un input dans la mosaïque complexe dans laquelle les décisions politiques émergent pour la planification, la modélisation, l'implantation et la perpétuation des programmes pour améliorer la condition humaine.» (Rossi & Freeman, 1985, p. 27.)

L'article de Dubois et Marceau (2005) développe encore davantage la réflexion sur les fondements théoriques des choix évaluatifs. Selon les auteurs, lorsque l'évaluation obéit à une démarche scientifique, ce sont «les théories qui guident la recherche empirique». L'évaluation est ainsi réalisée sur «la base de modèles qui orientent l'évaluateur dans son choix» (p. 12).

C'est bien dans une orientation de ce type que nous allons élaborer une procédure visant à fonder une évaluation en la situant dans son contexte, par rapport à ses référents (théoriques, sociaux et humains), en fonction des projets et des dynamiques de ses acteurs et au moyen d'indicateurs représentatifs de cet ensemble.

L'importance des référents

Selon la majorité des auteurs, comme nous l'avons évoqué en première partie, évaluer consisterait à confronter deux types de données: des données qui sont de l'ordre du fait observé (référés) et des données qui sont de l'ordre des éléments de comparaison choisis pour attribuer du sens au fait (référents, données «faisant référence», objectifs, normes, etc.). L'élucidation des référents amènerait en fait à (re)construire un «référentiel d'évaluation» (*cf.* plus loin) qui témoignerait d'une «certaine conception» de l'objet évalué (Figari, 1993, pp. 38-40; 1994, p. 180).

La référentialisation, comme processus de modélisation des référés et des référents

Dans l'évaluation « formelle », les « référés » sont la plupart du temps produits, comme nous l'avons souligné plus haut. C'est-à-dire que l'objet à évaluer n'apparaît pas à l'état brut : au-delà de l'institution ou du programme à observer, le but de l'évaluation est précisé par un problème ou une question particulière. S'agit-il d'évaluer des faits ? des comportements ? les résultats d'une politique ou d'un projet ? des compétences ? un mode de gouvernance ou de gestion sociale ? Généralement, l'angle de questionnement ne fait pas partie du mandat d'évaluation ou, en tous cas, il demande à être précisé, reformulé, affiné : le référé est donc à expliciter sinon à (re)construire.

Allons plus loin : si le référé est un « dispositif », c'est-à-dire un mécanisme qu'on construit pour produire un résultat précis (un programme, un projet, etc.), il constitue un système complexe par nature et ne peut être soumis directement à une évaluation sans passer par une modélisation. Il s'agit donc, pour évaluer, de se représenter son objet en le « modélisant ».

La modélisation que nous expérimentons et que nous utiliserons dans cette présentation, tout en s'inspirant du modèle bien connu de Stufflebeam (CIPP), propose de considérer ses éléments non comme linéaires mais comme itératifs et interactifs. Ainsi, le « programme-dispositif » peut être représenté par un schéma triangulaire en trois dimensions pouvant agir et interférer les uns avec les autres (Figari, 1994, p. 57) : les *Intrants* (éléments contextuels et finalités), les *Constructions* des acteurs (processus), les *Produits* – ou *extrants* – (effets, résultats, évolutions).

Modélisation d'un « programme - dispositif »

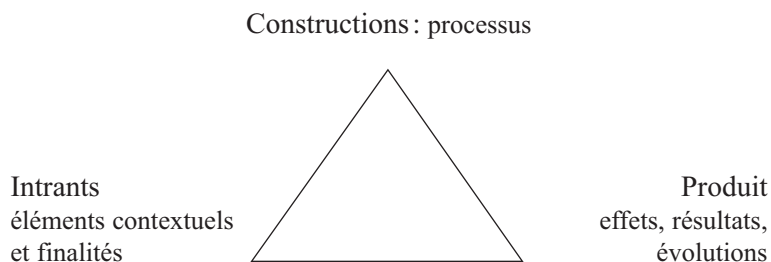


Figure 1 : ***Modélisation d'un « programme-dispositif ».***

Évaluer le «programme-dispositif» revient alors à étudier différentes comparaisons possibles :

- soit entre **I** (intrants) matérialisé par exemple par les objectifs d'un programme et **C** (construction) illustré par exemple par les stratégies développées par le programme pour étudier la pertinence du dispositif,
- soit entre **I** (intrants) et **P** (produit) si l'on souhaite étudier la rentabilité du dispositif,
- soit entre **C** et **P** si l'on évalue l'efficacité d'une stratégie, d'une méthode ou d'un outil.

Ce qui revient, pour chaque question, à utiliser toujours la même procédure : établir une comparaison entre un référé et un référent («référé C» versus «référé I» = pertinence; «référé P» versus «référé I» = rentabilité; «référé P» versus «référé C» = efficacité). On voit alors que le statut de «référé» n'est pas nécessairement lié à son contenu mais peut être réélaboré par l'évaluateur.

La «référentialisation» est d'abord, comme on vient de le voir, une procédure de modélisation de l'objet à évaluer (le modèle «ICP» pour représenter le dispositif) et de l'évaluation (confrontation référé/référent). Elle commence par établir la représentation que se construisent les évaluateurs du «système de références» qui est destiné à interpréter à conférer du sens aux résultats.

La fonction des critères et des indicateurs

Les référents sont traduits en «critères» destinés à guider l'observation. «Comme les données de l'évaluation, les critères ou les normes de jugement n'existent pas à l'état naturel; ils sont toujours le résultat de processus spécifiques conduisant à leur élaboration» (Barbier, 1985). Ces processus ne seraient explicités que lors des évaluations «instituées» ou «formelles», pour reprendre le terme de Nadeau. Le critère est généralement défini comme un principe qui permet de distinguer le vrai du faux, de juger, d'estimer. Il serait discriminant. Ce serait en fait l'interface entre des «visions du monde» et des «indices» (Ardoino & Berger, 1989, p. 215).

L'indicateur, comme son nom l'indique, «indiquerait» quelque chose : il constituerait le signe (observable et mesurable) de la manifestation du critère (par exemple, l'indicateur de réduction des incidents critiques serait le signe de l'efficacité d'une méthode d'organisation). Il n'aurait pas de sens en lui-même, mais en référence à un critère (Figari, 1994, p. 110). C'est pourquoi la réélaboration des critères et des indicateurs en fonction des référents constitue la phase centrale de la «référentialisation».

La référentialisation comme méthodologie de l'évaluation

Nous allons, dans un premier temps, rappeler, d'une manière générale, les phases du protocole de référentialisation utilisé dans un certain nombre de contextes et, dans un deuxième temps, illustrer ce protocole par le résumé d'un exemple dans lequel il a été mis en œuvre.

Qu'attend-t-on d'une méthode d'évaluation de «programme-dispositif»? De manière générale, comme dans toute opération d'évaluation, on peut espérer qu'elle puisse présenter les caractéristiques suivantes :

- définir l'objet à évaluer (c'est-à-dire préciser les aspects de l'institution qui demandent à être examinés);
- le situer dans un contexte et en repérer les acteurs (c'est-à-dire établir une première série de variables indépendantes, d'ordre socio-économique par exemple, qui fournissent, dès l'abord, sinon une première explication du phénomène à observer, du moins une hypothèse de travail);
- en fixer les référents (c'est-à-dire attribuer à tels ou tels éléments extérieurs à l'objet une fonction de «comparants» par rapport auquel il sera mesuré);
- élaborer des critères et des indicateurs d'évaluation (c'est-à-dire des qualifiants de l'appréciation et des traces de leur existence ou de leur importance);
- renseigner les indicateurs avec des données (par le recours aux questionnements et aux observations);
- interpréter et restituer les résultats (c'est-à-dire élaborer une phase d'intégration des diagnostics et des recommandations en prévoyant les interactions avec les partenaires de l'évaluation).

Le tableau 1 (Figari et al., 2005) présente une vision synoptique d'une méthode d'évaluation en formulant, pour chaque étape de la procédure de référentialisation, des questions à résoudre.

Tableau 1
Vision synoptique d'une méthode d'évaluation

<i>Opération d'évaluation = élucidation du mandat (commande)</i>
Quelle origine? Dans quel cadre? Quel est le contexte et quelles sont les finalités de l'évaluation?
<i>Objet à évaluer</i>
Quoi évaluer? Quelles dimensions de l'objet va-t-on évaluer?
<i>Référents</i>
Par rapport à quoi va-t-on attribuer un sens à l'information? (lois, projets, contexte national et international, état de la concurrence, etc.)
<i>Critères</i>
Quels aspects représentatifs de l'objet à évaluer va-t-on choisir pour porter un jugement?
<i>Indicateurs</i>
Quels indices, quelles traces, quelles caractéristiques observables pour vérifier le degré d'atteinte du critère (existence, taux, relation, impact, etc.)?
<i>Items de recueil des données (outils)</i>
Quels seront, pour chaque indicateur, les descripteurs qui le renseigneront? (un indicateur pouvant être décrit par un ou plusieurs descripteurs) Quel outil construire avec ces items?
<i>Interprétation</i>
Quels croisements de données effectuer? Quelles significations attribuer aux résultats?
<i>Restitution</i>
Que dire aux différents acteurs de l'évaluation? Comment? Comment l'exploiter?

Tableau 2
Exemple d'évaluation d'un programme de formation de formateurs

Opération d'évaluation	Objet à évaluer	Référénts	Critères	Indicateurs
Évaluation d'un programme européen de formation de formateurs au Portugal	I- Structure I-1 éléments de la structure :	Pour l'ensemble des éléments à évaluer :	Complétude du produit :	Présence des éléments : • titre • sommaire • objectifs • mode d'emploi • apports de connaissances • activités pédagogiques • activités d'évaluation • ressources internes • ressources externes
= contrôle et validation des produits « technico-pédagogiques » destinés à la formation des adultes = Exemple du manuel d'accompagnement de la formation	I-2 organisation des données II- Présentation II-1 forme d'édit.	• respect des normes du projet (cahier des charges spécifique pour la production d'outils pédagogiques) • respect des normes d'édition • référence à l'expertise didactique des disciplines visées	Lisibilité du plan formel Respect des normes du projet ...	Présence des caractéristiques : • découpage précis entre parties et sous-parties • fonctions différentes entre parties et sous-parties • présentation aérée du sommaire ... Formes graphiques : • polices • taille de caract. • format • interlignes • rapport évident entre texte et image ...
	II-2 illustrations II-3 formes d'expression		Adéquation au contenu à illustr. ... Homogénéité entre les parties ... Adéquation à l'utilisation	• styles de langage semblables dans chaque partie • Poids, volume ... existence d'accessoires de consultation
	III- Utilisation III-1 caract. physiques III-2 modalités d'utilisation IV- Types de contenus Etc.		Facilité de consultation

Un exemple : application à l'évaluation d'un programme de formation de formateurs. Le cas de l'évaluation d'un support pédagogique : le manuel de formation

Cet exemple est extrait de l'évaluation d'un programme européen ADAPT de formation de formateurs d'adultes au Portugal² (2001). Dans cette partie de l'opération, il s'agissait de contrôler et de valider des supports pédagogiques élaborés par les formateurs et comprenant des manuels, des fiches, des outils multimédia, etc. Il n'est présenté ici que la trame (non complète) du référentiel d'évaluation élaboré à cette occasion et ayant servi de base à la construction des outils qui a suivi cette phase de «référentialisation».

On vient d'examiner un aperçu (partiel) de protocole de référentialisation préalable à une évaluation de «programme-dispositif». Cette procédure, très complexe et très rigoureuse, laisse entrevoir sa face cachée : le travail de construction appuyé sur une préenquête et une discussion des choix de référents et de critères avec les acteurs de l'évaluation du programme (institutionnels, professionnels, universitaires), travail qui n'a pas été décrit ici. Le but de cette présentation était de donner à voir une modélisation possible de ce type d'évaluation de manière à illustrer l'enjeu de cet exposé qui était de contribuer à une confrontation théorique et méthodologique sur «l'évaluation de programme».

Conclusion

Par cette présentation, nous avons posé quelques jalons pour tenter de tirer des théorisations de l'évaluation quelques éléments qui permettent de mieux comprendre les pratiques, dans une perspective de comparaison internationale que nous n'avons fait qu'esquisser ici et qui serait à poursuivre dans d'autres recherches. Nous avons enfin tenté d'illustrer la façon dont une méthode d'évaluation, ici la «référentialisation», permettrait de prendre en compte ces éléments dans un but pratique. De même, les éléments théoriques que nous avons retenus comme constituant des «invariants» de l'activité évaluative seraient à confronter à l'étude des activités réelles des évaluateurs afin de tester leur validité.

Toutefois, malgré ces limites, nous pensons que ce cadre d'analyse permet de faire avancer les réflexions sur les pratiques d'évaluation, les rapports entre les théories qui les décrivent ou qui les fondent et leur application dans des cas concrets. C'est donc à un dialogue entre traditions françaises de l'évaluation (avec leurs influences européennes) et traditions québécoises que nous invitons ici.

NOTES

1. L'expression de «programme-dispositif» est volontairement employée ici dans le but de désigner l'objet qu'il s'agit d'étudier, tant du point de vue québécois que du point de vue français.
2. Programme européen ADAPT. Projet Delfim – (2001). Rapport GREIF (Figari & Gautier, 2001). Lisbonne: APAF.

RÉFÉRENCES

- Ardoino, J., & Berger, G. (1989). *D'une évaluation en miettes à une évaluation en actes*. ANDSHA, Matrice.
- Barbier, J. M. (1985). *L'évaluation en formation*. Paris: PUF.
- Chen, T. H. (1990). *Theory-driven Evaluations*. Newbury Park, CA: Sage.
- Cranton, P. (2001). Interpretative and Critical Evaluation. *New Directions for Teaching and Learning*, 88, 87-97. San Francisco: Jossey-Bass.
- CSE (1996). *Petit guide de l'évaluation des politiques publiques*, la documentation Française, Paris.
- Dubois, N., & Marceau, R. (2005). Un état des lieux théoriques de l'évaluation. *The Canadian Journal of Program Evaluation*, 20(1), 1-36.
- Figari, G. (1993). Quel système de références pour évaluer une mésostructure éducative? In D. Massé (éd.), *L'évaluation institutionnelle en milieu scolaire* (pp. 27-54). Université de Sherbrooke: Éditions du CRP.
- Figari, G. (1994). *Évaluer: Quel référentiel?*. Bruxelles: De Boeck Université.
- Figari, G., et al. (2005). Extrait des documents de l'Atelier «les fondements méthodologiques de l'évaluation». Poitiers: ESEN.
- Figari, G., & Gautier, G. (2001). *Évaluation des ressources technico-pédagogiques*. Rapport GREIF au Programme européen ADAPT (Projet Delfim). Lisbonne: APAF.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.
- Hadji, C. (1989). *L'évaluation, règles du jeu*. Paris: ESF.
- Hadji, C. (1997). *L'évaluation démystifiée*. Paris: ESF.
- Hurteau, M., & Houle, S. (2005). Évaluation de programme et recherche évaluative: des activités distinctes. *Mesure et évaluation en éducation*, 28(3), 83-95.
- Hurteau, M., & Duval, J. (2005). *Le rôle des standards dans une démarche évaluative*. Communication à la 27^e session d'étude de l'ADMEE, L'instrumentation: une évidence de l'évaluation... pas toujours évidente!, 24 et 25 novembre 2005, Québec.
- MEANS (1999). *Évaluer les programmes socio-économiques*. Communautés européennes, 6 volumes.
- Nadeau, M. A. (1988). *L'évaluation de programme, théorie et pratique*. Québec: Les Presses de l'Université Laval.
- Patton, M. Q. (1980). *Qualitative evaluation methods*. Beverly Hills, CA: Sage.
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: a systematic approach* (3^e édition). Beverly Hills: Sage publications.

- Secrétariat du Conseil du trésor (2002). *L'évaluation de programme. Document destiné aux dirigeants et dirigeantes de ministères et d'organismes* [www.tresor.gouv.qc.ca].
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of Program Evaluation Theories of Practice*. Sage publications.
- Stake, R. E. (1980). Program evaluation, particularly responsive evaluation. In W. B. Dockrell & D. Hamilton (éds). *Rethinking educational research* (pp. 72-87). London : Hodder and Stoughton.
- Stufflebeam, D., al. (1980). *L'évaluation en éducation et la prise de décision*. Ottawa: NHP.
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction, Syllabus for Education 360*. Chicago : University of Chicago Press.
- Younès, N. (2006). *L'effet évaluation des enseignements universitaires par les étudiants*. Thèse de doctorat, Université Pierre-Mendès-France, Grenoble.