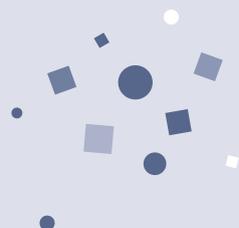


Série Notes méthodologiques

exPost
ExPost



Les évaluations sont-elles utiles ?

Revue de littérature sur « connaissances et décisions »

Jocelyne Delarue, Jean-David Naudet, Véronique Sauvat

Agence Française de Développement

Cette série Expost «Notes méthodologiques» vise à accueillir les travaux relevant de l'analyse des pratiques, des méthodes et des utilisations de l'évaluation des politiques et interventions de développement.

Avertissement. Les analyses et conclusions de ce document sont formulées sous la responsabilité de leurs auteurs. Elles ne reflètent pas nécessairement le point de vue officiel de l'Agence Française de Développement ou de ses institutions partenaires.

Directeur de la publication : Jean-Michel SEVERINO

Directeur de la rédaction : Jean-David NAUDET

ISSN : 1958-590X

Dépôt légal : janvier 2009

Mise en page : Solange Münzer

SOMMAIRE

1.	Introduction	4
2.	L'utilisation de la recherche : cadrage théorique	5
	2.1. Les différents modèles	5
	2.2. Des preuves d'utilisation peu nombreuses	6
	2.3. Les facteurs déterminants : les interfaces	9
	2.4. Les facteurs déterminants : le contexte et la nature des résultats produits	11
	2.5. Les facteurs déterminants : processus politiques et apprentissage organisationnel	13
3.	L'utilisation de l'évaluation	18
	3.1. L'évaluation : un domaine spécifique	18
	3.2. L'utilisation des connaissances issues de l'évaluation	19
	3.3. L'influence de l'évaluation	22
	3.4. Les facteurs déterminants de l'utilisation/influence	24
	3.5. L'indépendance	25
	3.6. L'objectivité	26
	3.7. La faisabilité	29
4.	Conclusion	32
	Bibliographie	34

1. Introduction

L'utilisation de l'évaluation pour améliorer les politiques publiques est au cœur de nombreux débats actuels dans le secteur de l'aide au développement, notamment en lien avec la mobilisation sur l'efficacité de l'aide et le développement des évaluations d'impact. Une vaste littérature s'intéresse à cette question, s'inspirant pour le cadre théorique de travaux plus généraux sur l'utilisation de la recherche en sciences sociales ou, plus étroitement, des résultats avérés de cette recherche (*evidence use*) pour l'amélioration des politiques publiques.

Cette libre revue de littérature entend tirer de sources choisies des éléments pour alimenter ce débat d'actualité.

Dans la section 2, le cadrage théorique de l'utilisation de la recherche dans la décision sera décrit à partir de la littérature existante sur les différents modèles d'utilisation et les facteurs qui l'influencent.

L'évaluation fera plus directement l'objet de la section suivante. Les spécificités de l'évaluation vis-à-vis de la recherche seront abordées, avant un examen des modèles d'utilisation et des facteurs déterminants, en fonction de ces caractéristiques particulières.

Une courte conclusion tentera de dégager les principaux enseignements que les auteurs retirent de l'examen de cette littérature, au regard de leur propre expérience.

2. L'utilisation de la recherche : cadrage théorique

Cette section présente les principaux résultats de la littérature sur l'utilisation de la recherche et souligne quelques aspects ou questions peu traités par ces travaux. Les principaux modèles d'utilisation font l'objet de la première sous-section. Cette architecture conceptuelle permet de mieux appréhender ce qui apparaît à certains analystes, et plus encore aux observateurs occasionnels, comme une sous-utilisation de la recherche – point qui sera abordé dans la deuxième sous-section.

Les trois sous-sections suivantes traitent tour à tour des trois « domaines » qui apparaissent comme des facteurs déterminants de cette utilisation : les interfaces entre recherche et décision ; le contexte et la nature de la recherche ; la nature des processus de décision et du fonctionnement des organisations.

Cette discussion amène à aborder la littérature sur de vastes domaines contigus à celui de l'utilisation de la recherche, tels que les différentes formes de connaissance, les institutions apprenantes ou les modèles d'élaboration de la décision politique. Ces questions complexes sont traitées ici schématiquement par le biais de synthèses de littérature déjà existantes et dans le seul but d'éclairer la discussion sur l'utilisation de la recherche et de la connaissance.

2.1. Les différents modèles

Carol Weiss a joué un rôle pionnier en élaborant une typologie des modes d'utilisation de la recherche. Elle proposait en 1979 sept schémas distincts (cf. encadré 1).

Encadré 1

Pluralité des modes d'utilisation de la recherche

La typologie proposée par Weiss en 1979 concerne l'utilisation de la recherche en sciences sociales dans la sphère des politiques publiques. Elle distingue sept modèles, dont les cinq premiers supposent que la recherche porte sur des sujets qui intéressent déjà les décideurs :

- *le modèle tiré par la connaissance* : le simple fait que la recherche a produit un résultat conduit à sa prise en compte et à son utilisation. Ce modèle tiré des sciences naturelles a peu d'exemples en sciences sociales ;
- *le modèle de résolution des problèmes* : ce modèle part de la formulation d'un problème à résoudre et d'une décision politique à prendre, la recherche venant aider à identifier et sélectionner des réponses (soit déjà existantes, soit issues de nouvelles recherches). C'est le modèle théorique d'utilisation de la recherche qui prévaut, mais il ne se vérifie pourtant que rarement ;
- *le modèle interactif* : dans ce modèle, la recherche n'est qu'un élément d'un faisceau d'informations à la disposition du décideur. Les consultations sont donc itératives et réciproques pour fonder une connaissance utilisable par le décideur ;
- *le modèle politique* : les résultats de la recherche sont utilisés pour justifier une position politique déjà prise, en ignorant les résultats qui la contredisent ;
- *le modèle tactique* : le fait de mener une recherche sur un sujet est utilisé en soi pour démontrer la responsabilité des décideurs qui la commandent, retarder la prise de décisions non désirées ou justifier la prise de décisions impopulaires ;
- *le modèle de l'éclairage progressif* : dans ce modèle, des résultats issus de travaux de recherche divers pénètrent progressivement le public informé et changent sa vision, éventuellement sur un sujet complètement nouveau ;

– la recherche comme partie intégrante du projet intellectuel de la société : la recherche et la politique sont nourries par les courants de pensée contemporains et s'influencent mutuellement.

De nombreux auteurs ont proposé depuis leur propre typologie plus ou moins inspirée du schéma initial de Weiss ou tenté de synthétiser les travaux antérieurs (Nutley et al., 2003 ; Lavis et al., 2003 ; et Hanney et al., 2003 pour citer quelques exemples récents).

Bien que chacune des propositions ait sa propre spécificité, un consensus met en évidence trois modèles principaux :

- un modèle dit « instrumental » ou « rationnel », qui se manifeste par une influence directe d'un résultat de recherche sur une décision politique – le processus pouvant être tiré par la connaissance (une production de connaissance aboutit spontanément à une utilisation dans les politiques) ou par la résolution d'un problème (face à un problème, des décideurs commandent une recherche et en utilisent les résultats) ;
- un modèle dit « conceptuel » ou « incrémental » (ou encore d'enlightenment), où l'influence de la recherche se fait de manière indirecte et/ou cumulative, en apportant une connaissance qui modifie la façon dont les décideurs considèrent un problème et envisagent ses solutions (Porter et Hicks, 1995, p. 3) ;
- un modèle dit « tactique » ou « symbolique » (ou encore politique) dans lequel les décideurs utilisent, commandent ou sélectionnent des résultats de recherche pour défendre des politiques existantes ou supporter des décisions déjà prises et des changements souhaités.

2.2. Des preuves d'utilisation peu nombreuses

Lorsque l'on parle sans plus de précision de l'utilisation de la recherche dans la décision publique, le modèle instrumental vient spontanément à l'esprit. Le constat général de la littérature sur le sujet est alors celui d'une faible utilisation de la recherche.

« For a long time social scientists were discouraged by the seeming failure of social science to alter policy. They expected one study and its recommendations to have a direct and immediate effect on policy action – to change it from A to B by the strength of its scientific findings. Once in a while, they found a “nugget”, an exemplary case where social science did in fact change policy outcomes. But usually it seemed as though social science was ignored. » (Weiss, 1986, p. 280)

La littérature diagnostique en général à ce propos un fossé entre chercheurs et décideurs.

« However, if we focus on the direct, instrumental use of research, a recurrent feature of the literature is the identification of a research-practice gap, in which evidence of what works in a particular field is not translated appropriately into actual practice. » (Nutley et al., 2003)

Ce fossé se traduit dans la pratique par la coexistence de deux mondes à part, disposant de leurs propres culture, agendas¹ et langages, et s'ignorant largement l'un l'autre.

« Une partie importante de la production scientifique se fait dans des organismes spécifiques, possédant leur "culture" propre et leurs règles : prétention à la maîtrise des questions qu'elle entend poser à la réalité, construction de théories testables et non pas de théories "faisables", refus des déterminations "prescriptives" ou de la vulgarisation réductrice, débats importants mais peu ouverts aux non-spécialistes. » (Jacob, 2000)

¹ Les chercheurs *chronophiles* face aux décideurs *chronophobes*, comme le dit de manière amusante Julio Frank (1992) in Hanney (2004), ce qui peut être illustré par cet extrait de Weiss (1986) : "Members of the House of Representatives spend about 11 minutes a day reading – and their greatest complaint is that they do not have time to study and analyse. When they do read something other than staff memos, it tends to be summaries of pending bills, crucial mail, and newspaper headlines. Just about nobody in high office reads social science journals".

Ce constat d'une faible utilisation, en particulier instrumentale, de la recherche est à l'origine de l'*evidence-based movement* (encadré 2), qui plaide de manière influente en faveur d'une utilisation rationnelle des faits probants issus de la recherche. Ce mouvement met en avant la qualité scientifique de la recherche, en sciences sociales en particulier, ainsi que le manque d'ordre qui peut prévaloir dans celle-ci.

Encadré 2

Evidence-based movement

L'*evidence-based movement* s'est d'abord développé dans le domaine de la recherche médicale : l'*evidence-based medicine* consiste à baser les décisions cliniques, non seulement sur les connaissances théoriques, le jugement et l'expérience – qui sont les principales composantes de la médecine traditionnelle – mais également sur des « preuves » scientifiques.

Dans le courant des années 1990, un certain nombre de chercheurs et de décideurs ont souhaité élargir cette démarche aux politiques publiques et aux recherches du champ social. Il s'agit alors de promouvoir la définition de politiques et de pratiques fondées sur les résultats issus de recherches rigoureuses. L'*evidence-based movement* postule qu'un lien de causalité peut être établi entre interventions et effets attendus sur une population : la recherche sert alors non seulement à décrire ce qui se passe mais également à tester des dispositifs et à prescrire des politiques, dans une vision instrumentale. Pour y parvenir, ce mouvement soutient en outre une amélioration de la qualité scientifique de la recherche et, en particulier, sa capacité à fournir des résultats probants de nature causale sur les politiques par l'utilisation de méthodologies répondant à cet objectif (notamment les démarches expérimentales - ou quasi expérimentales - et les « revues systématiques de recherches » - ou méta-analyses).

Le réseau de référence de l'*evidence-based medicine* est The Cochrane Collaboration, organisation internationale créée en 1993. The Campbell Collaboration, mise en place en 2000 dans le prolongement de celle-ci, touche trois domaines principaux : les politiques sociales, les politiques pénales et judiciaires, l'éducation. Au Royaume-Uni, l'Evidence network a été créé en 2000 pour couvrir tout l'éventail des activités nécessaires à la prise en compte des résultats de la recherche dans les politiques du pays.

Sources : Nutley et al. (2003) et Rey (2006).

L'utilisation conceptuelle de la recherche est certainement plus courante, mais elle est par nature extrêmement difficile à établir. Le lien entre connaissance et décision y est diffus et indirect et donc peu détectable, sinon par des investigations approfondies des mécanismes et au cours d'entretiens avec des acteurs ayant participé à une décision.

« *It is one thing to talk about knowledge creep and enlightenment; it is another to describe how social science travels to policymaking arenas, which channels are activated, which personal links matter, what personal experience or characteristics make policy makers more receptive, what situations trigger attentiveness* ». (Weiss, 1986, p. 278)

Au niveau agrégé, le lien peut être évident : ainsi, à la suite de l'accumulation des connaissances sur le changement climatique, de multiples décisions concrètes découlent de l'évolution de la perception qui s'est opérée chez les décideurs. Mais au niveau micro d'une décision ou d'une recherche particulière, les liens de cause à effet deviennent difficilement traçables. De ce fait, le critère d'utilisation ne peut pas être un guide pour définir la pertinence ou la qualité d'un travail scientifique.

« *Il n'y a pas une franche dichotomie entre utilisation et non-utilisation ; plus exactement, l'utilisation du savoir de la recherche se construit sur un changement progressif de la conceptualisation* ». (Carden, 2004, p. 155)

Dans ce modèle, le lien indirect entre décideurs et chercheurs se fait via un « réservoir de connaissances » que les chercheurs abondent, et qui doit éclairer (*enlighten*) la perception des problèmes par les décideurs et la modifier de façon incrémentale. Le caractère cumulatif de la connaissance est alors un facteur déterminant de l'utilité de la recherche, de même que la pertinence des sujets traités.

De l'avis des auteurs concernés, l'utilisation tactique des résultats de la recherche est probablement importante. Mais, même si certains exemples sont parfois évoqués dans la littérature, ce sujet semble curieusement avoir été très peu étu-

dié². Sans doute le modèle tactique est-il en bonne partie imbriqué dans les autres modèles, en particulier le modèle conceptuel.

En effet, la recherche est loin de produire uniquement des résultats convergents ou cumulatifs : ainsi, il n'y a pas un seul « corps de preuves » monolithique. Davantage de recherche ajoute au contraire à la complexité du sujet traité et le rend moins lisible pour les décideurs (Weiss, 1979). La question de la synthèse des différentes données du réservoir de connaissances apparaît donc fondamentale : sans ce nécessaire travail, beaucoup de résultats vitaux pourraient ne jamais parvenir à la connaissance des décideurs ou leur parvenir trop tard. Cela exige en particulier de trier entre les différentes recherches existantes : cette étape peut-elle se réaliser sur des critères objectifs et partagés (pertinence, qualité, etc.) ? On voit là se dessiner la subtile frontière entre utilisation conceptuelle et utilisation tactique. L'utilisation sélective de certains résultats de la recherche est souvent un reproche porté aux décideurs ou aux institutions politiques (encadré 3).

La frontière entre utilisation tactique et mauvaise utilisation (*misutilization*) de la recherche est également complexe et peu explorée. On évoque parfois un rôle de démonstration : la recherche apporte des données probantes sur ce qui marche. Cette fonction de démonstration peut se révéler indispensable pour convaincre les parties prenantes (parfois l'opinion publique) du bien-fondé d'une décision et la rendre ainsi possible – par exemple, l'extension d'une expérience pilote réussie. Cette fonction constitue potentiellement une utilisation positive de la recherche et peut amener à des utilisations instrumentales, s'il y a reproduction des expériences dont les résultats ont été démontrés.

Mais le modèle tactique est également souvent proche de ces efforts de démonstration. La sélection des exemples que l'on cherche à démontrer, puis à reproduire, permet de véhiculer par la même occasion un message positif sur l'efficacité des politiques. N'est-ce pas le cas de l'aide au développement où les efforts importants consacrés à illustrer les *success stories* ressortissent à la fois du souci de reproduire des expériences heureuses mais aussi de communiquer plus globalement sur l'efficacité des politiques d'aide au développement ?

Encadré 3

L'évaluation de la recherche conduite par la Banque mondiale

La Banque mondiale est une institution à la fois très influente dans les politiques et grande productrice de recherche et de connaissance. Cette double caractéristique lui donne une force incontestée dans l'utilisation des résultats de la recherche dans la décision, mais lui fournit également l'occasion de développer des utilisations tactiques, comme le remarquent les évaluateurs de la recherche de l'institution.

« *Bank researchers have also done extremely visible work on globalization, on aid effectiveness, and on growth and poverty. In many ways they have been the leaders on these issues. But the panel had substantial criticisms of the way that this research was used to proselytize on behalf of Bank policy, often without taking a balanced view of the evidence, and without expressing appropriate skepticism. Internal research that was favorable to Bank positions was given great prominence, and unfavorable research ignored. There were similar criticisms of the Bank's work on pensions, which produced a great deal that was useful, but where balance was lost in favor of advocacy. In these cases, we believe that there was a serious failure of the checks and balances that should separate advocacy and research. The panel endorses the right of the Bank to strongly defend and advocate its own policies. But when the Bank leadership selectively appeals to relatively new and untested research as hard evidence that its preferred policies work, it lends unwarranted confidence to the Banks' prescriptions. Placing fragile selected new research results on a pedestal invites later recrimination that undermines the credibility and usefulness of all Bank research* » (Banerjee et al., 2006).

Ce paragraphe, extrait de la synthèse de l'évaluation, doit naturellement être remis dans le contexte d'un ensemble faisant état de jugements très positifs par ailleurs. Il va sans dire qu'il s'applique probablement également à d'autres institutions comparables quant au lien entre la recherche qu'elles produisent directement et l'utilisation qu'elles en font.

² Citons néanmoins Neilson (2001), qui analyse cette question dans le champ du développement à partir, notamment, des travaux de Stone et de Krastev et des concepts respectifs de *perceived influence* et *faking influence*. Neilson pose la question : « *whether policy influence should always be construed as a positive development outcome* ».

2.3 Les facteurs déterminants : les interfaces

Le fossé entre décideurs et chercheurs apparaît, nous l'avons vu, comme le premier obstacle à l'utilisation de la recherche. La mise en évidence de ce phénomène est souvent attribuée à Caplan (1979) et son article pionnier "*The Two-Communities Theory and Knowledge Utilization*". Le franchissement de ce fossé est une véritable difficulté car souvent, les incitations pour les uns et les autres manquent pour surmonter les différences de culture, de centres d'intérêt ou de gestion du temps.

« The structure of incentives within the academic community has also driven a wedge between social scientists and policymakers. These incentives attach greater weight to knowledge building as against policy-forming research; to authoritativeness rather than usefulness; to the pursuit of rigor as against relevance; to the values of scientific independence as against the virtues of policy involvement; and to understanding rather than action. » (Booth, 1988)

Ces deux communautés ont également des perceptions distinctes de ce que constitue la connaissance (Booth, 1988 ; Neilson, 2001 ; Kennedy, 1983, pp. 193-211).

« Working knowledge is the organized body of knowledge that administrators and policy-makers use spontaneously and routinely in the context of their work. It includes the entire array of beliefs, assumptions, interests and experiences that influence the behavior of individuals at work. It also includes social science knowledge... Working knowledge often has greater cumulative influence on policies and practices than does the evidence that is selectively brought to formal decision points. » (Kennedy, 1983 cité par Garcia, 2008)

Une première piste pour réduire cette distance entre communautés est naturellement de provoquer la multiplication des contacts directs, ce qui revient en général à impliquer les décideurs et les praticiens dans des processus de recherche orientés vers la « résolution de problèmes ».

Innvaer et al. (2002) étudiant l'utilisation de 24 analyses d'impact dans le domaine de la santé font ainsi apparaître les contacts personnels entre chercheurs et décideurs comme

premier facteur déterminant. Les deuxième et troisième facteurs d'utilisation sont la pertinence et l'opportunité (*timeliness*) de la recherche ainsi que la nature claire et succincte des recommandations.

La « recherche-action » constitue dans ce sens un processus particulier de relation entre chercheur et décideur, qui se caractérise par une collaboration prolongée sur un sujet donné et une validation collective des objectifs de la recherche. Mais cette recherche de contacts directs, tournée vers l'utilisation instrumentale, possède sans doute ses limites. Elle s'adresse davantage à des praticiens et à des microdécisions qu'à des décideurs politiques. En outre, son succès n'est pas garanti.

« It does not follow from our data, that an alliance of social scientists and policy makers is the panacea which will produce relevant research and allow translation of the results of scholarly analysis into terms of practical politics. The notion that more and better contact may result in improved understanding and greater utilization may be true, but there are also conditions where familiarity might well breed contempt rather than admiration. The need for reciprocal relations between knowledge producers and knowledge users in policy-making is clear, but the problem of achieving effective interaction of this sort necessarily involves value and ideological dimensions as well as technical ones. » (Caplan, 1979)

Au-delà des contacts personnels, l'implication commune de décideurs et de chercheurs tout au long du cycle de la production de résultats de recherche est également mise en avant par Hanney (2004) qui distingue trois étapes-clés de ce cycle de recherche : (i) la définition des priorités de la recherche et des besoins de connaissance ; (ii) la commande de recherche ; et (iii) la diffusion et le transfert de connaissance.

Dans une enquête sur l'utilisation de la recherche dans les politiques de santé des pays en développement, Hennink et Stephenson (2004) retiennent notamment qu'un facteur déterminant d'utilisation de la recherche est le fait qu'elle ait été commandée par des décideurs.

« Differences in the dissemination process, however, exist between research that is commissioned (by a donor or a government agency) and non-commissioned research.

Respondents identified that with commissioned research there is a direct channel of communication between the researcher and the end-user, which facilitates the dissemination of the final research outputs. » (Hennink et Stephenson, 2004)

La mise en avant de cette communication directe montre que c'est l'utilisation instrumentale qui est plutôt considérée par les répondants.

De manière plus large, et notamment en ce qui concerne l'utilisation conceptuelle de la recherche, l'interface ne peut se ramener à des contacts directs entre chercheurs et décideurs. Ce rôle doit être pris en charge par des groupes ou des institutions spécialisés, légitimes du point de vue scientifique et capables de nourrir le dialogue politique à partir du réservoir de connaissances existant. Ces interfaces sont de différentes natures. Stone et al. (2001) retiennent ainsi quatre types de « communautés de connaissances » influentes dans les processus politiques : les réseaux thématiques (issue networks), les communautés scientifiques (*epistemic communities*), les communautés sectorielles (*policy communities*) et les coalitions de plaidoyer (*advocacy coalitions*). Les *think tanks* peuvent illustrer ces interfaces, souvent dans le rôle d'animateurs de ces différents groupes et réseaux. Le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) est un exemple d'interface entre chercheurs et décideurs dans le domaine du climat. L'*evidence-based movement* présenté dans l'encadré 2 en est un autre.

Des personnalités se prêtent également, de plus en plus fréquemment, à des rôles de ce type. Une synthèse sélective de connaissances disponibles portée par une personnalité reconnue et influente auprès des décideurs semble avoir souvent un impact très important (Walter et al., 2003, cité par Nutley et al., 2003). Le risque *a contrario* avec ce mode de diffusion est éventuellement la sélection de résultats en raison de leur caractère spectaculaire plutôt que de la rigueur avec laquelle ils ont été obtenus (Weiss, 1979).

En résumé, ces interfaces remplissent trois fonctions essentielles pour faciliter l'utilisation de la recherche : (i) elles contribuent à orienter la recherche autour de la résolution de problèmes posés aux décideurs ; (ii) elles font connaître la

recherche dans les milieux des décideurs et aident à en rendre les résultats légitimes et pertinents pour ces derniers ; et (iii) elles mélangent différents types de connaissances et font en particulier entrer des résultats de sciences sociales dans les savoirs pratiques et professionnels.

Encadré 4

La déforestation en Guinée : une lecture sélective des données probantes

La vision des experts et des politiques sur la zone de contact forêt-savane dans la préfecture de Kissidougou en Guinée est révélatrice de la façon dont le discours « scientifique » est parfois induit par des croyances établies et vient légitimer un discours politique bien ancré.

La vision prédominante sur le milieu naturel de cette région considère depuis toujours que les îlots forestiers autour des villages sont des reliques d'une forêt originelle dense, qui aurait auparavant couvert toute la région. Ce diagnostic de « savanisation » de la région a été répété par les administrateurs coloniaux, par les différents scientifiques qui se sont succédé sur cette région et par les experts internationaux. Il a été la base de politiques répressives des pratiques locales, allant jusqu'à instituer la peine de mort pour ceux qui déclencheraient des feux de brousse, pour tenter d'enrayer ce qui était perçu comme une destruction du milieu. De nombreux et coûteux programmes de coopération internationale sont jusqu'à présent mis en œuvre pour reboiser le milieu.

Or, la comparaison des photographies aériennes et satellites aux différentes époques, ainsi qu'une étude scrupuleuse des pratiques locales faite par Fairhead et Leach (1996) démontrent au contraire que l'activité humaine tend à faire progresser le milieu forestier sur la formation naturelle d'origine, qui est la savane. Les observations menées jusqu'à présent par les experts extérieurs étaient simplement conditionnées par des théories de changement de la végétation qui ne laissaient aucune place à la compréhension du réel.

Malgré l'étude de Fairhead et Leach et de nombreux travaux sur les zones de transition forêt-savane en Afrique de l'ouest, le discours alarmiste sur la déforestation de la zone de Kissidougou n'a jusqu'à présent guère évolué.

Les interfaces sont indispensables à une utilisation large de la recherche dans la décision. Mais les groupes en question sont également plus ou moins parties prenantes du processus politique. Ils défendent eux-mêmes des valeurs, des conceptions scientifiques, voire des intérêts et peuvent être amenés à une lecture sélective du réservoir de connaissances (encadré 4). Leur rôle important pourrait laisser penser que les modèles conceptuels et tactiques ne constituent pas des alternatives, mais que bien des processus réels d'utilisation de la recherche empruntent conjointement à ces deux modèles.

2.4. Les facteurs déterminants : le contexte et la nature des résultats produits

Le contexte dans lequel les résultats sont produits est un déterminant majeur de leur utilisation. Ce contexte tient à la nature des processus de décision (section 2.5) mais dépend également étroitement du moment où les résultats sont produits et de la façon dont ils sont présentés. L'opportunité (*timeliness*) et la faisabilité apparaissent comme des facteurs cruciaux d'utilisation de la connaissance dans la décision.

Les situations où il est légitime de prendre des décisions importantes – fin de cycle d'une politique, changement d'équipe, changement de mandat, etc. – sont par nature favorables à l'utilisation de connaissances qui viendraient à point nommé. Cela peut être le cas des situations de crise qui appellent des processus décisionnels distinctifs³. Plus largement, Weiss propose une liste de ce type de situations d'opportunités.

« The occasions that give rise to search of evidence are (1) new issues, something the policymaker has not dealt with before and therefore needs to get an orientation to, (2) big questions with important or expansive consequences, (3) issues on which they feel inadequately prepared, and (4) situations where their judgement may be challenged and they want authoritative support » (Weiss, 1986, p. 276)

³ Les auteurs voient dans ces situations de crise des éléments qui peuvent être favorables ou défavorables à l'utilisation de connaissances, selon les cas (Neilson, 2001).

Kingdon (1984) parle de « fenêtres politiques » pour qualifier ces opportunités décisionnelles propices à l'utilisation des connaissances – en insistant sur leur caractère souvent fugace.

« As I see, people who are trying to advocate change are like surfers waiting for the big wave. You get out there, you have to be ready to go, you have to be ready to paddle. If you're not ready to paddle when the big wave comes along, you're not going to ride in it. » (Kingdon, 1984, cité par Neilson, 2001, p. 32)

La question de la faisabilité englobe le contexte et la nature des résultats produits par la recherche. Le fait d'être orienté vers des questions opérationnelles ou politiques, de parvenir à des résultats probants et de mener à des recommandations à l'intention des décideurs, sont des éléments qui favorisent nettement l'utilisation. Cela concerne en particulier l'utilisation instrumentale de la recherche : plus la recherche s'inscrit dans une optique de « résolution de problème », qu'elle soit ou non commandée par les décideurs, plus elle sera facilement utilisée directement dans la mise en œuvre des politiques.

L'ampleur du changement qui découle des résultats obtenus est également un élément de la faisabilité.

« The instrumental use of research is in fact quite rare. It is most likely where the research findings are non-controversial, require only limited change and will be implemented within a supportive environment: in other words, when they do not upset the status quo. » (Weiss, 1998)

Cette dimension fait ressortir un autre point essentiel présent en filigrane dans cette littérature : la distinction entre utilisation de la recherche dans les pratiques et dans les politiques⁴. Une partie importante de la littérature analysée est issue du secteur médical où les connaissances utiles pour la pratique et les politiques médicales sont d'une nature différente. Néanmoins, une distinction du même ordre se révèle pertinente entre les décisions courantes de nature opérationnelle, qui ont trait au perfectionnement de l'existant (amélioration ou « maintenance » des politiques, renforcement des pratiques) et les déci-

⁴ Distinction que l'on retrouve dans l'*evidence-based movement* qui se scinde en *evidence-based practice* et *evidence-based policy* (et, plus récemment, en *evidence-based management*).

sions radicales relevant du politique, qui impliquent des changements en profondeur (abandon ou transformation radicale d'une politique, modification de l'allocation des ressources, etc.). La littérature relève que l'utilisation de la recherche, en particulier l'utilisation instrumentale, s'observe plutôt dans le domaine des décisions courantes.

Lorsque les résultats de la recherche sont invoqués par les parties prenantes à la source d'une décision radicale, il n'est pas rare qu'il s'agisse d'une utilisation tactique de connaissance en tant qu'argumentaire à l'appui de décisions déjà prises.

Dernier aspect relatif à la nature des résultats produits : la qualité des résultats et de la recherche qui y a conduit est-elle un facteur déterminant de son utilisation ? Cette question n'a pas de réponse simple et paraît curieusement peu étudiée (Neilson, 2001, p. 39). On peut cependant retenir deux positions potentiellement divergentes sur ce point.

La première met en avant l'exigence de rigueur au titre de la validité et de la crédibilité des résultats obtenus, tout en soulignant qu'il est du devoir de la communauté scientifique de veiller à cette rigueur afin de limiter les usages sélectifs et les mauvais usages de la connaissance que les décideurs peuvent être tentés de faire.

« The credibility of research can not be taken for granted. Certain practices are essential to maintaining the public stature of knowledge producers. Some research is more rigorous, professional and scholarly, adhering to recognized standards of peer review. Such standards need to be cultivated and protected as policy-makers and others usually require policy research and analysis produced in a professional context. On other words, they want research findings that help legitimate policy and these come from recognized institutions and experts. » (Stone et al., 2001)

C'est l'un des fondements de l'*evidence-based movement* que de revendiquer une forte exigence de rigueur afin d'apporter aux décideurs la connaissance la plus crédible. Les méta-analyses et les synthèses – l'une des tâches prioritaires de ce mouvement scientifique – se basent sur des critères de rigueur, dits objectifs, d'inclusion et d'exclusion des recherches.

La définition de critères partagés de rigueur et de qualité est une première difficulté à laquelle il faut faire face dans cette optique. Pour certains auteurs, cette lecture sélective de la recherche aboutit à ne retenir qu'un nombre limité de travaux par conformité à des critères trop formels (MacLure, 2005).

La seconde difficulté est qu'une approche « élitiste » de la qualité de la recherche, même définie de manière consensuelle, conduit à « chasser » d'autres formes de connaissances, notamment des connaissances pratiques et des savoirs professionnels, moins rigoureusement établis mais utiles aux décideurs notamment du fait de leur contextualisation et de leur proximité. Cette « concurrence » entre connaissance scientifique et connaissance pratique peut prendre des formes extrêmement polémiques et clivantes. C'est le cas autour de l'*evidence-based movement* : l'encadré 5, tiré d'un livre de Paul Krugman (1997), l'illustre de manière plus amusante et néanmoins significative.

L'hégémonisme de la seule connaissance légitimée par les communautés scientifiques est potentiellement aveuglant, en ce qu'il revient à ignorer, et parfois mépriser, les savoirs locaux ou l'expérience de terrain.

« Wood et al. (1998) found that health professionals do not simply abstract scientific research but collaborate in discussions and engage in work practices which actively interpret its local validity and value. There is no such thing as the body of evidence: evidence is a contested domain and is in a constant state of becoming. Research is rarely self-evident to the practitioner but varies according to the context in which it is received. » (Nutley et al., 2003, encadré 10)

Ce dilemme entre rigueur et proximité est illustré par Neilson (2001), à l'intérieur du champ scientifique, en posant la question de l'utilisation par les décideurs des pays en développement de la recherche provenant de leurs propres instituts.

« Research produced by developing country scholars is often considered to be of poor quality. For the most part, this is due to the limited amount of funding developing country governments allocate to research... As a result, policy makers in these countries often utilize research results or research

Encadré 5**The Evolution of European Ignorance about Africa**

A friend of mine who combines a professional interest in Africa with a hobby of collecting antique maps has written a fascinating paper called “*The Evolution of European Ignorance about Africa*”. The paper describes how European maps of the African continent evolved from the 15th to the 19th centuries.

You might have supposed that the process would have been more or less linear: as European knowledge of the continent advanced, the maps would have shown both increasing accuracy and increasing levels of detail. But that’s not what happened. In the 15th century, maps of Africa were, of course, quite inaccurate about distances, coastlines, and so on. They did, however, contain quite a lot of information about the interior, based essentially on second- or third-hand travellers’ reports. Thus the maps showed Timbuktu, the River Niger, and so forth. Admittedly, they also contained quite a lot of untrue information, like regions inhabited by men with their mouths in their stomachs. Still, in the early 15th century Africa on maps was a filled space.

Over time, the art of mapmaking and the quality of information used to make maps got steadily better. The coastline of Africa was first explored, then plotted with growing accuracy, and by the 18th century that coastline was shown in a manner essentially indistinguishable from that of modern maps. Cities and peoples along the coast were also shown with great fidelity.

On the other hand, the interior emptied out. The weird mythical creatures were gone, but so were the real cities and rivers. In a way, Europeans had become more ignorant about Africa than they had been before.

It should be obvious what happened: the improvement in the art of mapmaking raised the standard for what was considered valid data. Second-hand reports of the form “six days south of the end of the desert you encounter a vast river flowing from east to west” were no longer something you would use to draw your map. Only features of the landscape that had been visited by reliable informants equipped with sextants and compasses now qualified. And so the crowded if confused continental interior of the old maps became “darkest Africa”, an empty space.

Source : Krugman (1997).

consultants originating from industrialized countries, since this research is perceived to be of better quality and therefore a more credible base for decisions. The question remains, however, as to whether or not there is some trade-off between producing high quality research by outsiders or producing indigenous research that may be of lower quality and, therefore, may not have any impact on the policymaking process. However, the quality of the research, although important, is only one aspect of its likelihood to influence policymakers.» (Neilson, 2001, p. 41)

2.5. Les facteurs déterminants : processus politiques et apprentissage organisationnel

La question de l’utilisation de la recherche a été traitée jusqu’à du point de vue de la production de connaissances (l’offre en quelque sorte). Le point de vue peut être renversé en se plaçant du côté de la prise de décision. Plusieurs vastes pans de littérature méritent alors d’être abordés : l’analyse des mécanismes de la décision politique, les modèles cognitifs de la rationalité des décideurs et l’analyse des organisations. Ces sujets sont tour à tour effleurés uniquement sous l’angle du rôle de la connaissance. En effet, ces analyses traitent souvent de la connaissance comme un ensemble, sans distinguer clairement la particularité de la connaissance scientifique ou de la connaissance issue de la recherche.

Les processus de décision politique ont fait l’objet de nombreux travaux, principalement issus des sciences politiques, qui proposent une assez grande diversité de modèles et de cadres conceptuels. Neilson (2001) identifie dans la littérature six modèles principaux de prise de décision politique, dont les caractéristiques sont synthétisées dans l’encadré 6 page suivante.

Encadré 6**Les modèles de décision politique**Les modèles rationalistes

– **linear**: This model ... *“outlines policymaking as a problem-solving process which is rational, balanced, objective and analytical. In the model, decisions are made in a series of sequential phases, starting with the identification of a problem or issue, and ending with a series of activities to solve or deal with it”* (Sutton, 1999). Ce modèle, qui désagrège le processus politique en phases successives, a été opérationnalisé dans les pays en développement par les donateurs : *“the linear model has led donor agencies to support substantial efforts to strengthen policy analysis in developing countries in the expectation that good analysis will translate into good decision making and into good policy”* (Grindle et Thomas, 1990).

– **incremental**: *This model focuses on the decision making aspect of the policy process and presents it as one which takes small incremental steps towards policy change. It is useful to explain small changes in current policies but is unable to explain major changes or crises management.*

– **interactive**: *The interactive model views policy reform as a process, one in which interested parties can exert pressure for change at many points* (Grindle et Thomas, 1990). *This model focuses on the implementation phase of the policy process. Understanding the field, the actors involved and the context in which the process occurs are key.*

Les modèles politiques

– **political networks**: *« public policy results from conflict, bargaining, and coalition formation among of potentially large number of societal groups organized to protect or advance particular interests common to their members »* (Grindle, 1997). *This model better illustrates the complexities of decision making.*

– **agenda setting**: *This model explores the policy making process by looking at why and how some subjects and/or alternatives are considered for the policy agenda while others are not. This model analyses the meeting point between problems, solutions and policies. It emphasizes the role of activists and political entrepreneurs.*

– **policy transfer**: *« Policy transfer refers to a process by which actors borrow policies developed in one setting to develop programmes and politics in another. »* (Dolowitz et Marsh, 1996, pp. 343-57)

Source : Neilson (2001).

Ces modèles peuvent être rattachés à deux familles ou deux « camps », qui s'opposent sur la manière dont la recherche influence le processus politique : le camp rationaliste et le camp politique. Les modèles rationalistes, en particulier le linéaire et l'incrémental, font clairement de la recherche et des connaissances la « matière première » de la prise de décision et, en ce sens, leur armature conceptuelle fait directement écho aux modèles d'utilisation de la recherche explorés plus haut.

« A “rationalist” point of view is that new research can directly prompt policy change. The “political camp” on the other hand assume that various external factors play a key part both in defining the question that a research project tackles and in influencing the impact of the answers on policy. » (Philpott, 1999)

Les modèles rationalistes ont été critiqués pour leur description peu réaliste du processus de décision politique. Plus précisément, on retrouve dans ces critiques la distinction déjà opérée précédemment entre les décisions relatives à la maintenance des politiques – pour lesquelles les modèles rationalistes peuvent être pertinents – et les décisions visant un changement de politique – où les modèles politiques seraient prédominants.

Ces derniers font intervenir des acteurs variés et des jeux complexes entre ces acteurs. Ils s'appliquent à identifier les groupes/structures concernés par le processus de décision : les élites et les hauts fonctionnaires, mais aussi les groupes d'intérêt, la société civile, les « communautés de connaissance » (*think tanks*, réseaux, etc.) déjà mentionnés dans l'analyse des interfaces. Les connaissances issues de la recherche, qu'elles soient empiriques ou plus fondamentales, ne sont qu'un des facteurs susceptibles d'influencer la prise de décision, à côté (en concurrence avec) des systèmes de croyance, des intérêts, des rapports de force.

« The policymaking process is a political process, with the basic aim of reconciling interests in order to negotiate a consensus, not implementing logic and truth. » (Weiss, 1997)

Si les modèles insistent sur l'importance des spécificités des systèmes politiques pour le processus de décision, presque tous les travaux analysent le fonctionnement des systèmes

politiques démocratiques occidentaux. Le processus de prise de décision dans les pays en développement est très peu documenté⁵.

En France, l'analyse des politiques publiques s'est focalisée sur le fonctionnement de l'État. Le scepticisme par rapport à la rationalité de la décision publique (théorie du *public choice*) s'est traduit par l'analyse de la dimension cognitive de l'action publique et de la pluralité des logiques à l'œuvre.

« Deux perspectives sont privilégiées. La première, qui relève d'une sociologie des acteurs, s'intéresse aux producteurs des représentations qui orientent les décisions publiques. La seconde, qui relève d'une perspective sociogénétique, cherche à rendre compte du processus de construction de catégories au fondement des politiques publiques. » (Hassenteufel et Smith, 2002)

Cette incursion dans le processus cognitif des décideurs ouvre sur un autre domaine de la littérature portant sur les limites humaines de la rationalité décisionnelle.

« Les individus agissent selon une rationalité limitée : ils cherchent des solutions dans le voisinage immédiat de l'émergence des problèmes, essayent d'abord des modèles causaux simples, avant d'élargir progressivement leur quête si la connaissance produite ne rend pas des résultats satisfaisants. » (Cyert et March, 1963, cités par Baumard, 2002)

Cette rationalité limitée des individus tient autant à leur incapacité à percevoir l'ensemble des dimensions d'un environnement complexe (Cyert et March, 1963), qu'à leur difficulté à verbaliser et rationaliser l'ensemble de leurs connaissances. Nonaka et Takeuchi (1995) ont conceptualisé ce second phénomène en établissant une distinction entre connaissance explicite (formalisée, codifiée et stockée) et connaissance tacite (incorporée à son support humain, non formalisable). Les travaux ultérieurs ont établi l'importance de ces connaissances tacites dans la capacité d'innovation des entreprises (Hedlund, 1991).

La place des mécanismes rationnels dans la décision est également éclairée par l'ensemble des travaux ayant trait à l'analyse des organisations. On retrouve la difficulté d'isoler dans des processus complexes ce qui relève de l'utilisation de connaissances, et plus encore si l'on veut se limiter à des types particuliers de connaissance.

« Il est très difficile dans une organisation de distinguer ce qui appartient au domaine des croyances et des schémas collectifs, ce qui relève de mythes rationnels (qui peuvent avoir été empruntés aux discours circulants de la société au sens large), de ce qui pourrait être défini comme une connaissance en soi » (Baumard, 1995, pp. 49-57).

Ou encore :

« Dans l'urgence, et pour préserver les flux d'action dans lesquels ils sont socialement engagés, les managers dans l'organisation trouvent d'abord des solutions toutes prêtes (des répertoires comportementaux) et essayent de les raccorder aux problèmes qui émergent. Ainsi les organisations, en tant que génératrices d'action, sont plus souvent dans un processus d'invention à la volée ou de raccrochage aux idéologies ou discours circulants, que dans un processus d'épuration et d'élaboration objective de connaissances. » (Baumard, 2002)

La question de l'utilisation de la connaissance est au cœur des processus d'apprentissage dans les organisations. Une modélisation très influente de ces processus a été proposée par Argyris et Schön (1978) et développée depuis dans de nombreux travaux. Ces auteurs distinguent l'apprentissage en simple boucle, relatif à l'amélioration de l'existant, de l'apprentissage en double boucle qui implique des remises en cause paradigmatiques (encadré 7 page suivante). Les organisations apprenantes se caractériseraient par l'existence de cet apprentissage en double boucle.

⁵ Le modèle interactif de Grindle et Thomas (1990), appliqué aux périodes de réformes structurelles des années 1980, constitue une exception.

Encadré 7

Le modèle des boucles d'apprentissage

Argyris et Schön (1978) ont cherché à dépasser la vision classique, linéaire et hiérarchique de la décision et du management dans l'entreprise, en s'intéressant aux processus d'apprentissage dans les organisations. Partant d'une approche cognitive centrée sur l'individu, ces auteurs mettent en évidence les comportements défensifs qui incitent les agents à ne pas remettre en cause les valeurs dominantes de leur entreprise, quand bien même ils ne les partagent pas ou en connaissent les failles.

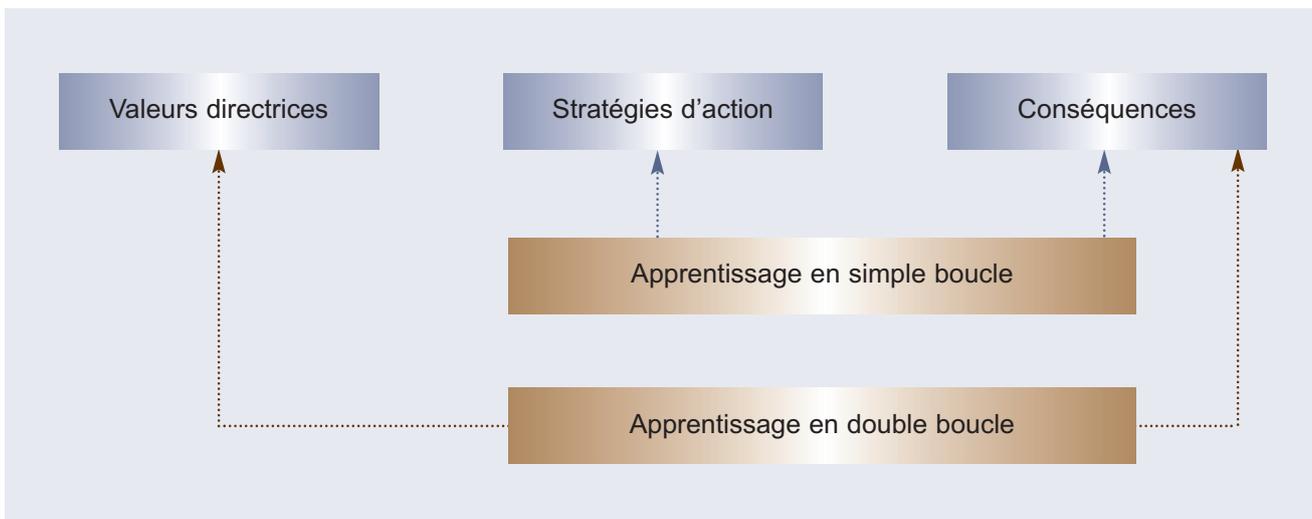
Ils distinguent deux modèles de boucles d'apprentissage :

- *le modèle en simple boucle (single-loop learning)* met l'accent sur la détection et la correction des erreurs, dans le cadre d'un système de valeurs donné. Argyris et Schön illustrent ce processus en prenant

l'image du thermostat : *"single-loop learning is like a thermostat that learns when it is too hot or too cold and turns the heat on or off"*. L'effet de *feedback* joue comme support d'innovation, limitée à des changements incrémentaux dans l'organisation ;

- *le modèle en double boucle (double-loop learning)* suppose la mise en question des valeurs directrices elles-mêmes. Ce processus entraîne le plus souvent des changements radicaux tels qu'une remise à plat complète du système, des réorientations de la stratégie de l'organisation, etc.

Cette lecture dichotomique des processus d'apprentissage, qui peuvent générer soit des décisions opérationnelles (d'amélioration de l'existant), soit des décisions radicales (de remise en cause de l'existant), fait écho aux modèles d'utilisation de la recherche (*problem solving* ou *enlightenment*). La séparation par niveaux de décision (décisions courantes contre décisions radicales) apparaît réellement comme un point fixe des réflexions sur l'utilisation de la connaissance, qui traverse l'ensemble des champs disciplinaires d'analyse.



L'utilisation des connaissances sous différents angles (acquisition, diffusion, interprétation, mémoire) est un sujet crucial de la littérature sur les organisations apprenantes. Cette dernière est cependant d'une nature assez spécifique : elle relève de l'analyse managériale autant que de la recherche et est largement tournée vers des buts prescriptifs. Elle est souvent plus « idéaliste » que descriptive.

« Learning organizations are more likely than other organizations to be adaptable and flexible, lack complacency, experi-

ment with and create new knowledge, rethink means and ends, and tap people's potential for learning as a strategic competitive advantage. » (Redding, 1997, cité dans Cors, 2003)

Les recommandations pour une meilleure utilisation des connaissances à partir de ces travaux varient toutefois selon les auteurs. Tout d'abord, l'acquisition de connaissances a un coût : les connaissances externes ne sont ni libres ni gratuites et, pour être en mesure de les exploiter, les organisations doivent construire et entretenir une capacité d'absorption (Bellon,

2002 ; Pénin, 2003). La production endogène de connaissances est elle-même coûteuse à mobiliser du fait de son caractère hétérogène et largement tacite (Foray, 1998). La mémorisation et le transfert d'une expertise exigent une reconstruction complète du savoir (Foray, 1993).

Nonaka et Takeuchi (1995) recommandent de multiplier les lieux et les occasions de sociabilité dans l'organisation, de manière à créer un environnement propice à l'émergence de connaissances tacites. D'autres auteurs préconisent de limiter la formalisation des connaissances en période d'exploration et d'innovation et de faire place aux compétences d'improvisation organisationnelle (Chédotel, 2005). Hatchuel (1999) met l'accent sur le « principe de non-séparabilité connaissances/reactions » et élabore un modèle d'interactions entre production de connaissances et action à l'appui d'une relecture des modalités de l'action collective, tandis que Cohendet et Llerena (1999) explorent en quoi le glissement vers une

conception de la firme comme processeur de connaissances invite à reconsidérer les structures de gouvernance et les mécanismes d'incitation internes.

Si la littérature n'est pas encore claire sur les modalités opérationnelles, le balayage des travaux met en évidence le fait que la nature et la gestion des connaissances sont devenues un enjeu pour les organisations – que ce soit sous l'angle de la compétitivité ou sous celui de la réduction de l'incertitude (Machlup, 1993). Certaines firmes adoptent ainsi – et de façon distincte de leurs besoins en termes de prise de décisions – une stratégie délibérée de codification, d'émission et de diffusion de leurs connaissances à l'extérieur, en vue d'établir leur légitimité sur un certain champ de compétences, d'être reconnues et d'intégrer des réseaux spécialisés dans ces domaines, de participer à l'élaboration de normes, de faire connaître ce dont elles disposent mais aussi ce dont elles ont besoin (Péning, 2003).

■

3. L'utilisation de l'évaluation

Cette section présente le cas particulier de l'utilisation de la connaissance issue de l'évaluation. Une première sous-section situe la particularité de l'évaluation en tant que connaissance, notamment vis-à-vis des connaissances issues de la recherche, abordées ci-avant.

Les deux sous-sections suivantes analysent la littérature sur les différents modes par lesquels l'évaluation interagit avec les processus de décision. Les modes d'utilisation de l'évaluation sont examinés selon une approche historique, avant de traiter des mécanismes plus larges de l'influence de l'évaluation.

La quatrième sous-section présente brièvement les facteurs déterminants de l'utilisation/influence de l'évaluation, en complétant les éléments donnés en première partie et qui restent largement valables dans le cas de l'évaluation.

Enfin, les trois dernières sous-sections examinent tour à tour des questions plus spécifiques à l'évaluation, toujours du point de vue de son utilisation : l'indépendance, l'objectivité et la faisabilité.

3.1. L'évaluation : un domaine spécifique

Une ligne de partage entre une appréhension « positiviste » et une appréhension « constructiviste » de l'évaluation traverse la littérature théorique sur ce sujet. La nature de l'évaluation, ses modes d'utilisation et les facteurs déterminants de son impact sur les décisions sont abordés différemment selon ces deux perspectives.

Dans une approche plutôt positiviste, l'évaluation est vue comme une forme de recherche appliquée destinée à informer le jugement plutôt qu'à juger.

« Evaluation is a type of policy research, designed to help people make wise decisions about future programming. Evaluation does not aim to replace decision makers' experience and judgement, but rather offers systematic evidence that informs experience and judgement. » (Weiss, citée dans Alkin, 1980, p. 83)

Dans la même mouvance, Rossi et al. (2004) qualifient globalement l'évaluation de recherche appliquée et la distinguent de la recherche en général à partir de plusieurs spécificités que l'on peut résumer comme suit :

- **la qualité** : l'évaluation adopte une approche pragmatique de la qualité qui doit être « suffisante » pour répondre aux questions pratiques posées, alors que la recherche sera plutôt dans la poursuite de la meilleure qualité possible, et notamment de la meilleure méthodologie possible, sous réserve de faisabilité ;
- **l'initiative** : alors que la recherche répond le plus souvent à un intérêt endogène, qui la dirige vers les questions reconnues pertinentes au sein d'une communauté disciplinaire, l'évaluation est le plus souvent impulsée par une demande externe tournée vers la résolution de problèmes pratiques ou politiques ;
- **le champ** : alors que la recherche est le plus souvent cantonnée dans une seule discipline du point de vue des questions posées et des méthodes utilisées, l'évaluation tente de répondre à des questions faisant intervenir une multiplicité de connaissances et doit pouvoir passer aisément d'un champ à un autre ;
- **l'utilisation** : le public de la recherche est avant tout un public de pairs, la validation et l'utilisation se font essentiellement au sein d'une communauté disciplinaire ; à l'inverse,

l'évaluation est jugée par des commanditaires externes et est destinée à leur être utile ;

● **l'éthique** : présente dans la recherche, la question éthique apparaît néanmoins beaucoup plus importante dans le domaine de l'évaluation.

Les spécificités de l'évaluation relèvent de deux domaines : en interne, l'évaluation se caractérise par une vision pragmatique de la qualité, la mise à profit de connaissances diversifiées selon des méthodes propres et des champs d'étude larges ; en externe, elle se distingue par des processus ouverts et dirigés de l'extérieur de la profession. Ce dernier point est plus particulièrement mis en avant par Rossi et al. (2004).

« Two essential features of the context of evaluation must be recognized: the existence of multiple stakeholders and the related fact that evaluation is usually part of a political process. »

L'évaluation est en partie indissociable du processus de décision lui-même. À ce titre, certains auteurs voient une séparation plus nette entre recherche et évaluation.

« Social science research may be differentiated from evaluation on contextual or valuing dimensions. First, evaluation is typically defined as systematic inquiry to judge the merit, worth or significance of some entity... A second distinction is that despite remarkable overlap in methods with social science research, evaluation by definition is context-bound; knowledge is produced for a particular purpose for a particular set of users in a specific context. » (Cousins, 2006, p. 2)

On est renvoyé à la double nature de l'évaluation comme recherche appliquée d'une part, et comme jugement argumenté à l'usage des décideurs d'autre part. L'activité d'évaluation couvre un large spectre qui va de la simple notation d'un « objet » donné à des analyses scientifiques complexes. Au pôle « positiviste », on trouvera de vastes programmes de recherche destinés à mesurer l'impact d'un programme avec le souci d'une objectivité maximale. Au pôle « constructiviste », on peut imaginer un compte rendu de l'opinion spontanée des acteurs ou des bénéficiaires sur le déroulement et l'efficacité d'un programme, destiné à informer rapidement les décideurs.

3.2. L'utilisation des connaissances issues de l'évaluation

Jusqu'à la fin des années 1990, la littérature retient une typologie des modes d'utilisation de l'évaluation proche de celle qui vaut pour la recherche. L'évaluation est davantage considérée sous l'angle positiviste d'une forme de recherche appliquée et certains des théoriciens de l'utilisation de la recherche sont d'ailleurs des spécialistes de l'évaluation, en premier lieu Carol Weiss.

À titre de synthèse, Rossi et al. (2004) proposent dans un travail de référence sur l'évaluation une typologie en trois niveaux :

« First evaluators prize the direct (instrumental) utilization, that is, the documented and specific use of evaluation findings by decision makers and other stakeholders. Second, evaluators value conceptual utilization. As Rich (1977) defined it, conceptual utilization refers to the use of evaluations to influence thinking about issues in a general way. A third type of use is persuasive utilization, the enlisting of evaluation results in efforts to support or to refute political positions – in other words, to defend or attack the status quo. »

Les évaluateurs utilisent également les termes de rétroaction et d'apprentissage. On peut voir une correspondance approximative avec les concepts utilisés ci-dessus : la rétroaction relevant plutôt de l'utilisation directe et l'apprentissage plutôt de l'utilisation conceptuelle.

Si les concepts paraissent presque similaires pour l'évaluation et la recherche, les mots utilisés ne sont pas tout à fait les mêmes et sans doute, l'importance attribuée à chacun de ces modèles diffère.

Les spécificités de l'évaluation font que l'utilisation directe instrumentale y est particulièrement attendue : structuration du questionnement autour de la résolution de problèmes, processus pilotés directement par les décideurs, importance donnée aux recommandations avec parfois formalisation du suivi de leur utilisation, etc.

« Drawing on a survey of US evaluators, Preskill and Caracelli (1997) report that 99 percent of respondents agree that providing information for decision-making and improving program – the two classic forms of instrumental use – are motivations for doing an evaluation. » (Henry et Mark, 2004, p. 36)

Néanmoins, sans être en mesure de faire un quelconque bilan empirique de ce type d'utilisation, c'est toutefois un constat plutôt désabusé qui prédomine chez les spécialistes de l'évaluation. Cracknell (2000) évoque la rétroaction comme le « talon d'Achille » de l'évaluation. Weiss (citée dans Alkin, 1990) admet que « even the best and greatest evaluations only minimally affect how decisions get made ». Rossi et al. (2004) présentent un bilan plus équilibré.

« Disappointment about the extent of the utilization of evaluations apparently is due to their limited direct or instrumental use. It is clear that many evaluations initiated for their direct utility fell short of that mark. However, it is only in the past two decades that the extent of direct use has been systematically studied. These recent efforts challenge the previously held belief that evaluations do not have direct utility... In summary, there does seem to be a fair degree of instrumental utilization, although a pessimistic view on this point is still widely held among evaluators and potential consumers of evaluation. »

Furubo (1994 et 2005) donne une explication convaincante d'une possible double perception de l'utilisation directe de l'évaluation à partir de la distinction déjà évoquée entre décisions courantes opérationnelles et décisions radicales politiques. À partir de l'expérience suédoise, il convient que l'évaluation a conduit à des changements d'orientation politique seulement dans des cas exceptionnels. En revanche, les cas d'utilisation directe sont nombreux pour la « maintenance des politiques » et l'amélioration des opérations.

Weiss précise ce constat nuancé en définissant les conditions de l'utilisation instrumentale.

« At least instrumental use is common under three conditions: (i) if the implications of the findings are relatively non-controversial, neither provoking rifts in the organization nor running into conflicting interests, (ii) if the changes that are

implied are within the program's existing repertoire and are relatively small-scale, and (iii) if the environment of the program is relatively stable, without big changes in leadership, budget, types of clients served, or public support. There is a fourth situation that encourages use, too – when the program is in a crisis or paralysis, and nobody knows what to do. » (Weiss, 1998, p. 24)

Le deuxième modèle – l'utilisation conceptuelle – est moins renseigné, car relevant d'un processus de transmission la plupart du temps imperceptible. L'évaluation ne consiste pas seulement dans la mise au point de recommandations, elle fournit également des clés pour mieux comprendre les situations étudiées⁶ et en tirer des « leçons d'expérience » dont l'utilisation dans les actions futures est souvent diffuse.

« Evaluation evidence often comes into currency through professionals and academics and evaluators, and it influences networks of practicing professionals and policy works, infiltrate advocacy coalitions, alters policy paradigms, changes the policy agenda, and affects belief communities within institutions. » (Weiss, 1998)

Une notion clé pour ce type d'utilisation est certainement une nouvelle fois le caractère cumulatif ou non de la connaissance. Il n'est pas anodin de remarquer que bien que les concepts soient proches, on parle souvent de modèle « incrémental » en ce qui concerne la recherche (ce qui induit l'effet de l'accumulation de connaissance) et de modèle « conceptuel » pour l'évaluation (où la notion d'accumulation ne transparaît plus).

En effet, la connaissance évaluative est probablement moins apte à l'accumulation que la connaissance scientifique. La forte contextualisation des analyses, l'absence de communauté épistémologique, de validation par les pairs et de questions disciplinaires structurantes, mais aussi la faiblesse de procédure instituée de construction des connaissances (revues,

⁶ Ainsi, le baromètre 2008 de l'évaluation des politiques publiques, réalisé par Ernst and Young à partir d'une enquête auprès de plus de 1 000 répondants, rapporte qu'à la question « quelle est la contribution des rapports d'évaluation ? », la réponse la plus fréquente est « améliorer le niveau global de connaissance sur le secteur évalué » (44 % des réponses), devant la proposition de recommandations (39 %).

séminaires) font que la connaissance évaluative reste souvent parcellaire et éparpillée.

« *A number of commentators in a special issue of the American Journal of Evaluation, looking at the future of evaluation (Lipsey 2001, Mark 2001, Patton 2001 and Pawson et al. 2001) while coming from somewhat different perspectives, agree that we have developed little cumulative knowledge and tend to evaluate each program as if it is unique.* » (Perrin, 2005, p. 25)

L'accumulation de connaissances évaluatives est l'une des formes les plus courantes de ce que l'on nomme « capitalisation », c'est-à-dire l'accumulation progressive de connaissances opérationnelles utiles, issues de la mémoire d'une institution ou d'une communauté professionnelle.

C'est le sujet (même si le mot de capitalisation, peu utilisé en anglais, n'est pas repris) d'un ouvrage collectif synthétisant l'expérience d'évaluateurs de différents pays (Rist et Stame, 2005), qui défend l'utilité de l'évaluation dans les institutions à travers la création de « courants » d'information au profit des décideurs. Furubo soutient dans ce livre l'idée de banques de connaissance évaluative :

« *Thus, instead of the notion of a model in which evaluations are regarded as knowledge channeled directly to the political decision makers, an alternative way of looking at things is to use the analogy of a knowledge bank'. Different evaluations can contribute to this knowledge bank with information deposits.* » (Furubo, 2005, p. 159)

La simple accumulation d'informations n'est sans doute pas suffisante. De même que pour la recherche, le rôle des interfaces pour porter cette connaissance de façon utile, sélective et actualisée aux décideurs est sans doute capitale. Dans le même ouvrage, Perrin (2005) souligne l'importance de synthèses régulières et opérationnelles de la connaissance évaluative. Il distingue pour cela plusieurs types d'outils : évaluations de grappes d'actions similaires, synthèses d'évaluations, bilan thématique (*integrative review*). Il décrit également de manière détaillée à qui doivent être confiées ces synthèses et quel niveau d'importance doit leur être donné.

« *An ideal review/synthesizer should be an experienced person, familiar with various policy domains, rather than a neophyte or junior. Unfortunately too often review are approached in exactly the reverse way, resulting in reviews of questionable validity and limited use.* »

Comme pour la recherche, le troisième modèle – l'utilisation persuasive, parfois appelée également symbolique – semble peu étudié. Il est pourtant plus que probable qu'il s'agisse d'un mode d'utilisation plus courant que les autres.

Plusieurs auteurs relèvent par exemple un biais positif dans la publication (GAO, 1992 ; Pawson, 2006 ; Perrin, 2005), mais encore plus dans la citation d'évaluations : lorsque les évaluations renforcent une politique ou une position institutionnelle, elles sont davantage mobilisées et utilisées. Il est difficile d'éviter ce type de gestion sélective du « marché » de l'information par les décideurs et plus généralement par les « parties prenantes » d'une politique.

En ce qui concerne l'évaluation, Furubo (2005) met en garde sur la différence entre l'apprentissage et l'utilisation.

« *In a study of a political program it is said that the program has been well implemented and has been extremely successful in reaching its goals. The probability that the responsible minister will use this evaluation is very high! The audience will probably have to listen to the politician when, he or she boasts about the success of the program. So the evaluation will certainly be used. We can't exclude the possibility that the minister is also learning, but it doesn't follow from the form of utilization we have just hinted at. That's another story.* »

Le rôle de démonstration d'une évaluation mobilisée pour le soutien d'une politique ou d'une innovation peut avoir une utilité sociale élevée. Mais cela peut facilement donner lieu à un « mauvais usage » de l'évaluation. Ce concept de mauvais usage a également entraîné différentes sortes de classification dans la littérature (Cousins, 2004), qui restent cependant difficiles à analyser en pratique.

« *Several examples of misutilization have surfaced in the literature. These include: (a) the commissioning of evaluation*

for purely symbolic purposes; (b) the conscious subversion of the evaluation process by program practitioners; and (c) the purposeful non-use of high quality information... While misutilization is relatively easy to define in the abstract, studying it empirically is quite another matter. » (Shulha et Cousins, 1997, p. 18)

3.3. L'influence de l'évaluation

Progressivement, les analystes ont été amenés à considérer l'utilisation des résultats des évaluations comme un concept trop étroit pour embrasser la question de l'utilité sociale de l'évaluation.

Patton (1997) introduit une nouvelle catégorie d'usage de l'évaluation, qu'il lie au processus d'évaluation (*process use*) et non plus aux résultats de l'évaluation. Les acteurs de l'évaluation sont amenés à changer de compréhension, à rapprocher leur vue, à mieux comprendre les positions des autres parties, à reconsidérer la « théorie » de leur programme, etc., non pas à partir des enseignements de l'évaluation mais du fait de leur participation aux différentes étapes du processus de l'évaluation. L'apprentissage passe aussi par l'implication (*learning by involvement* – Cracknell, 2000).

« Process use is individual changes in thinking and behaviour, and program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process. » (Patton, 1997, p. 90)

Cet élargissement de l'utilité de l'évaluation est désormais reconnu et peut même être étendu, au-delà de la participation au processus, à des éléments de méthode ou de procédure de l'évaluation.

« Use of evaluation used to mean the use of results for making program decisions. Now we are aware of its larger dimensions. Many aspects of evaluation can be used – ideas and insights, signals that are sent by the very fact of evaluating, the processes of learning through cooperation with eva-

luation, the choice of outcome indicators that herald success, even the study design. Use encompasses a broad array of effects by multiple classes of users. » (Weiss, 1998)

Le rôle de « prévention » de l'évaluation, rarement mentionné, peut également être aisément imaginé, à défaut d'être observé.

« It is also possible that evaluations can be influential in stemming the momentum for change by preventing ineffective policies or programs from being funded and implemented. » (Henry, 2003, pp. 515-524)

La fonction de transparence et de redevabilité que remplit l'évaluation, peu traitée par la littérature examinée, peut naturellement être rapprochée de ce rôle préventif. L'existence même du principe d'évaluation – et l'obligation associée de rendre compte – peut modifier le comportement des décideurs indépendamment d'une quelconque utilisation de l'évaluation.

Cet élargissement de l'utilité de l'évaluation a été conceptualisé dans un premier temps par Kirkhart (2000) qui propose d'abandonner le concept d'utilisation de l'évaluation au profit de celui d'influence de l'évaluation. Ce concept, plus large, comprend les effets découlant du processus et des résultats, ceux découlant des effets intentionnels et non intentionnels et, enfin, des effets directs et immédiats comme des effets indirects et de long terme. On est sans doute proche de la notion d'impact de l'évaluation, c'est-à-dire de l'ensemble des changements, tous champs confondus, qui sont attribuables à « l'événement » évaluation. Kirkhart propose de passer de la question *“how are the results of evaluation study used”* à la question *“how and to what extent does evaluation shape, affect, support and change persons and systems”*.

Henry et Mark (2003) estiment également que se concentrer sur l'utilisation de l'évaluation fait perdre de vue le réel objectif de l'évaluation, qui est l'amélioration de l'efficacité des politiques.

« Use is inadequate as a motivation for evaluation because it fails to provide a moral compass... And falls short of recognizing and promoting the ultimate purpose of evaluation as social betterment. »

Les auteurs constructivistes ont également proposé de redéfinir le lien entre évaluation et décision. Ils s'appuient pour cela sur une perspective argumentative considérant l'évaluation comme une série d'arguments destinés à pénétrer dans la sphère où s'élaborent les décisions, puis à y être débattus, acceptés ou réfutés au même titre que d'autres arguments émanant de différents registres (idéologies, intérêts, normes institutionnelles, etc. (cf. encadré 8). On peut retenir de ces développements récents que le concept élargi d'influence de

l'évaluation donne lieu à des schémas logiques de plus en plus complexes. Les cadres conceptuels s'approchent peut-être de la diversité des voies par lesquelles l'évaluation affecte réellement les décisions, et surtout les conditions de vie des bénéficiaires des politiques. Mais on risque d'achopper sur un problème que les évaluateurs connaissent bien : celui de l'attribution. Comment discerner en pratique, dans l'entremêlement des chaînes causales envisagées, les effets relevant de l'évaluation proprement dite ?

Encadré 8

Le processus d'utilisation de l'évaluation

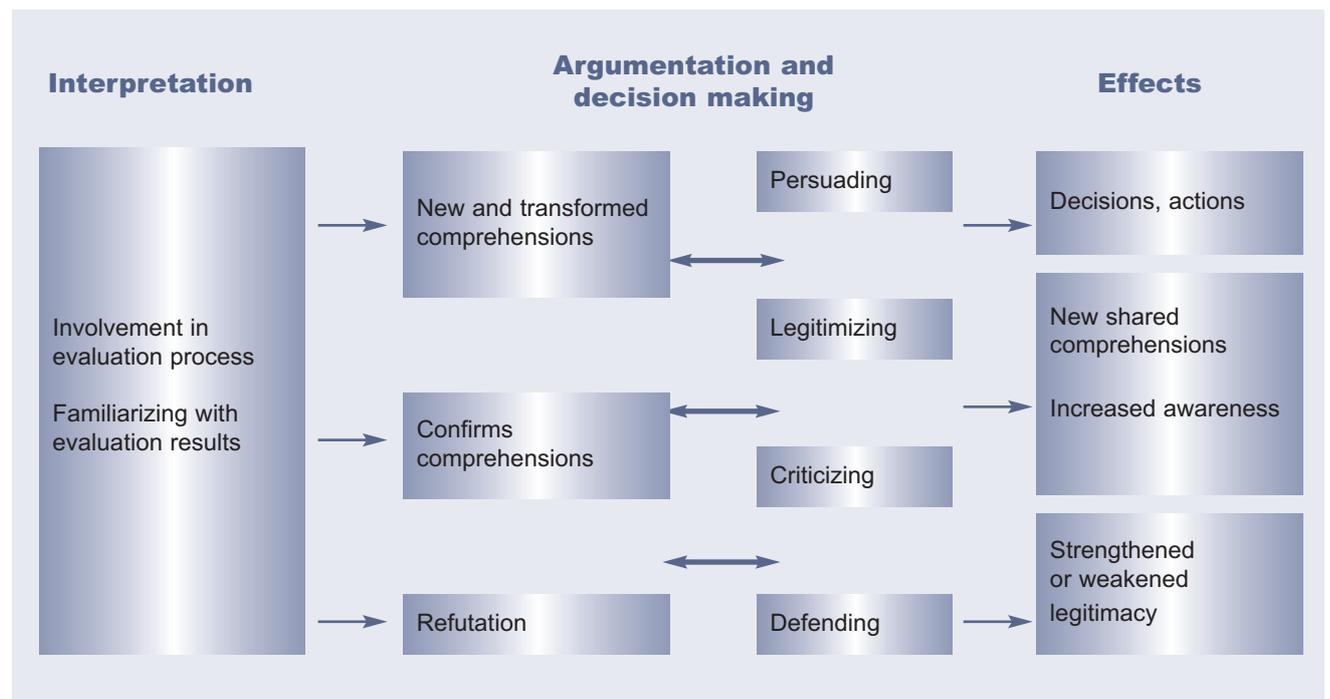
Valovirta (2002, pp. 60-80) conceptualise l'utilisation de l'évaluation en trois étapes.

L'interprétation peut venir d'une participation à l'évaluation ou d'une lecture du résultat. L'utilisateur examine la qualité de la recherche au regard de ses connaissances préalables et de ses attentes. Il examine dans le même temps l'utilité de l'information, c'est-à-dire ses implications pour une action possible et les défis qui en découleraient. Il en tire une compréhension nouvelle, confirme ses connaissances préalables ou réfute les nouveaux résultats.

L'étape d'argumentation et de prise de décision vient ensuite ; c'est un processus interactif où les résultats sont débattus, reconstruits et positionnés dans un contexte social précis. Elle combine la tentative de convaincre à un niveau cognitif, puis de persuader les autres d'agir.

Ces interactions peuvent aboutir à des décisions et à des actions, ou à une nouvelle compréhension partagée du problème. Elles peuvent déboucher également sur un renforcement ou un affaiblissement de la légitimité de l'institution.

Source : Valovirta (2002).



3.4 Les facteurs déterminants de l'utilisation/influence

La plupart des auteurs qui ont travaillé sur ces thèmes ont proposé une lecture des facteurs déterminants de l'utilisation/influence de l'évaluation. Dans une synthèse très complè-

te de la littérature sur ce thème, Garcia (2008) en propose une large liste, reprise dans l'encadré 9.

Une discussion sur l'ensemble de ces facteurs dépasse notre ambition. Beaucoup reprennent et élargissent les facteurs examinés en première partie, qui restent valables pour l'évaluation comme ils l'étaient pour l'utilisation de la recherche.

Encadré 9

Facteurs d'utilisation et d'influence

Users and use characteristics

Types of use to which the user puts evaluation; professional prestige and reputation of the user; numerosity – ability to comprehend numbers, statistical summaries, and other quantitative analyses; philosophical orientation (e.g. liberal, conservative); political ambition; working knowledge; open-mindedness, reflective practitioner stance; commitment to issues, ways of doing business, acting on evaluation; extent of change required for adoption; characteristics of the change/innovation suggested; operating orientation of the individual user; legal requirements (e.g. requirements to evaluate, need to apply scientifically validated interventions when publicly funded); propensity to take risks; intention to take a decision (or not); communication styles; primary intended user or broker/mediator; mindshare – amount of time to focus on a priority; expected tenure in role (i.e. temporary or permanent)

The evaluation itself

For whom the evaluation has been commissioned; purpose of evaluation – program improvement, summative judgment, knowledge generation, accountability; type of evaluation model (e.g. utilization-focused, empowerment, theory-based, dialogic, realistic etc.); relevance – the degree to which the evaluation actual attends to the needs of the user; type of information presented; report timeliness; technical quality; total elapsed time and cost; report readability, comprehensibility and interpretability; novelty of the findings and congruence with working knowledge/intuition; reaction elicited – positive/negative, consistency with audience expectations; what is measured, may be affected – what gets measured, gets done; competing findings/information; paucity/accumulation of prior research; scope of recommendations

The evaluator

Past reputation; credibility; organizational affiliation; independence – degree of cooptation; inside/outside of the organization and degree of interaction with the program; communication skills; personality characteristics; who compensates her/him and how much she/he is paid

Organizational (internal contextual) factors

Level within the organization; where the evaluation is done; where the use occurs; where and how analysis is completed; organizational constraints (e.g. bureaucratic requirements, sign-off, decision discretion); formal and informal structures; communication patterns, networks, and dissemination vehicles; time for consideration/reflection on evaluation; competition within the organization; history and age of the organization; existing policy – formal and informal; capacity – extent of technical knowledge, capacity for action/change; organizational learning mechanisms to interpret/make meaning and engage in corrective action

Environmental (external contextual) factors

Visibility of the evaluation to external forces; degree of politicization; partisan or multi-/pan-partisan interests; timing of the study – including in the program decision-making or policy cycle; inter-organizational competition; interest group and public expectations; size and political importance of constituencies; community characteristics; turbulence/complexity – economic, social, technical environment

Interpersonal and interaction characteristics

Interaction from planning phases; developing shared understanding; user involvement and advocacy of findings/recommendations; personal association, commitments and identity with evaluation; user turnover; relationship between the evaluator and decision-maker; degree of control over release and communication of information; information processing styles; "excessive evaluation anxiety"; "jointly negotiated decision-making and meaning making".

Source : Garcia (2008).

On s'arrêtera sur trois débats qui concernent plus spécifiquement l'évaluation : l'indépendance, l'objectivité et la faisabilité. Ces trois questions sont rapidement abordées ci-après, toujours à partir de la littérature existante sur l'utilisation/influence de l'évaluation.

3.5. L'indépendance

Il existe de nombreux modes d'organisation de l'évaluation des politiques publiques. L'institutionnalisation de l'évaluation est cependant peu abordée dans la littérature sur l'utilisation/influence de l'évaluation.

Dans le domaine du développement, la référence semble être celle du *separate unit model*. C'est le modèle de l'ensemble des organisations multilatérales, par exemple, pour lesquelles une unité indépendante d'évaluation rend directement compte au niveau politique (commissaire, parlement, *board*).

« *The separate unit model implies that responsibility for planning evaluations rests with a specialized unit which is managerially separated from the ordinary line positions of the organization responsible for intervention or programme being evaluated. Following this model the responsibility for planning evaluation is separated within the organisation.* » (Lundgren et al., 2003 cité par Foresti, 2007)

Il existe toutefois une tension classique entre la défense de la crédibilité de l'évaluation, fortement liée à son degré d'indépendance, et l'incitation à son utilisation, favorisée par l'implication des différents acteurs.

« *The best way that we know to date of encouraging use of evaluation is through involving potential users in defining the study and helping to interpret results, and through reporting results to them regularly while the study is in progress. As Huberman (1989) has recommended, such strategies may be even more successful if the evaluator maintains contact with users for a long period after the study ends and if the evaluator listens to what program people have to say about the meaning of evaluation findings and learns from them and with them about how to apply the findings in concrete situations. This*

kind of sustained interactivity' transforms one-way reporting into mutual learning. » (Weiss, 1998, p. 30)

Il est commun d'opposer les évaluations indépendantes aux évaluations collaboratives ou même participatives. Cette opposition est fondée sur le rôle des acteurs directs du programme à évaluer dans le processus d'évaluation : simples sources d'information dans l'optique indépendante, consultés et/ou associés dans les évaluations collaboratives aux différents stades de l'évaluation et enfin véritables parties prenantes dans les évaluations participatives, dont la version extrême est l'auto-évaluation.

Les évaluations indépendantes s'adressent au niveau de décision supérieur à celui du programme évalué (généralement le niveau politique) et aux types de décisions correspondantes (allocation des moyens, stratégies, politiques publiques, désignation des représentants publics).

C'est probablement avant tout à ce niveau que se constate le défaut d'utilisation instrumentale : rares sont les exemples de décisions politiques directement issues d'une évaluation. L'évaluation y est davantage considérée dans une logique argumentative, d'affaiblissement ou de renforcement du statu quo. Les utilisations conceptuelles, et surtout persuasives, sont sans doute des voies d'influence beaucoup plus fréquentes (Valovirta, 2002).

À l'inverse, les évaluations collaboratives sont avant tout destinées aux acteurs d'un programme (y compris le management direct du programme) et au niveau de décision qui les concerne : poursuite des financements, amélioration des programmes, aménagements institutionnels, etc. L'utilisation instrumentale y est sans doute plus fréquente mais non exclusive des autres modes. L'influence du *process* y est très importante puisque les acteurs sont associés à l'évaluation et peuvent être amenés à changer leur vue sur le programme ou à mieux comprendre les positions respectives au cours de l'évaluation.

Les évaluations participatives (Guba et Lincoln, 1989) se distinguent par l'étroite implication des acteurs dans le déroulement de l'évaluation (de la commande à la diffusion des résultats) mais aussi souvent par le rôle d'acteur donné aux « bénéficiaires », souvent ignorés dans les programmes d'évaluation

classique. Dans ce type d'évaluation, l'évaluateur est amené à s'engager dans des rôles de médiateur, de formateur et de catalyseur, ce qui le met en tension avec son rôle habituel d'investigateur (Cracknell, 2000). Toutes mettent en avant le respect du pluralisme des valeurs et de la diversité des connaissances.

Les évaluations collaboratives et participatives sont tendues vers l'enjeu de leur utilisation directe par les acteurs concernés. Dans ce sens, l'utilisation n'est pas un souhait mais l'objet même du travail et met en jeu la responsabilité de l'évaluateur. Cela conduit par exemple au concept d'« évaluation centrée sur l'utilisation » développé par Patton (1978).

En résumé, l'influence des évaluations peut emprunter plusieurs voies :

- une voie supérieure, en privilégiant des évaluations indépendantes organisées par des unités institutionnelles séparées ;
- une voie « intérieure », en recherchant des évaluations à la fois crédibles et collaboratives, destinées à un apprentissage collectif *in itineri* ;
- une voie inférieure, en favorisant la pratique et la culture de l'évaluation au niveau de chacun des acteurs.

La littérature ne permet pas d'instruire davantage cette question du mode opératoire afin d'institutionnaliser l'apprentissage des politiques publiques (*institutionalize policy-oriented learning*, Van der Knaap, 1995).

3.6. L'objectivité

Le débat quantitatif/qualitatif a souvent provoqué une fracture dans le monde de l'évaluation. Sous cette appellation réductrice, assimilée au seul usage de données quantifiées, se retrouve l'opposition entre approches positiviste et constructiviste : d'un côté, les évaluations dont l'objet est de mesurer « scientifiquement » le résultat d'une intervention précise ; de l'autre, des évaluations plus heuristiques, dont l'objet est de construire la meilleure connaissance possible sur des questions diverses à partir de l'ensemble de méthodes et de savoirs dis-

ponibles. Certains font remonter cette opposition à la construction des sciences sociales au XIXe siècle :

« *While early on the status quo tended to assert the positivist position underscoring the importance of quantification, experimentation and the establishment of generalized social laws, others proposed more heuristic approaches, suggesting that sociology should be about facilitating and understanding of meaningful actions.* » (Neilson, 2001)

La position positiviste s'appuie sur trois revendications : (i) ce sont les résultats qui comptent, le reste est subordonné ; (ii) l'objectivité tient à la qualité scientifique des mesures effectuées ; et (iii) la généralisation et la « découverte » de lois sociales va de pair avec un traitement rigoureux de la question de la causalité. Cette position s'appuie sur un instrument privilégié : les évaluations expérimentales, dites aussi « à assignation aléatoire » ou quasi expérimentales, qui construisent dans le champ social les conditions d'une expérience *in vivo*.

Ces points sont contestés par les tenants de l'approche constructiviste : (i) les *process*, les mécanismes d'apprentissage, les changements institutionnels sont, au-delà des résultats sur les bénéficiaires, des aspects cruciaux des politiques publiques ; (ii) la revendication d'objectivité des méthodes quantitatives est parfois excessive (faible prise en compte du moyen/long terme, méthodes complexes reposant sur de nombreuses conventions, résultats souvent réduits à un petit nombre d'indicateurs) ; et (iii) la causalité ne peut souvent être traitée que de manière probabiliste (Goldthorpe, 2001, cité par Iverson, 2003), car l'importance du contexte rend la généralisation hasardeuse. Plus fondamentalement, la contestation est portée au nom de la pluralité des questions, des connaissances, des méthodes, voire des points de vue et des valeurs.

Faut-il pourtant s'en tenir à cette opposition ? De nombreux auteurs soulignent, à l'instar de Cook (1997, p. 35), que ces évaluations sont complémentaires et non concurrentes : « *The case for qualitative methods does not depend on attacking the foundations of quantitative methods; it rests on their utility for answering important questions either when used alone or when used together with quantitative methods* ».

Ce débat dure depuis de nombreuses décennies et il n'est pas question de vouloir le résumer ici. Il est sans doute utile d'en noter la « violence »⁷. Celle-ci découle sans doute d'une hiérarchisation perçue de la valeur des évaluations issue de la position positiviste. De ce point de vue, les évaluations expérimentales sont considérées comme des références (*golden standard*) qui conduisent à des conclusions plus objectives et plus généralisables que les autres types d'évaluation. Les tenants de l'approche quantitative revendiquent régulièrement d'effectuer des synthèses sélectives des connaissances qui ne retiendraient que les résultats d'évaluation scientifiquement avérés.

Au-delà de cet « élitisme » méthodologique, la vigueur du débat est significative de nombreuses oppositions sous-jacentes à ces types d'évaluation : entre chercheurs et évaluateurs, entre disciplines qualitatives et quantitatives, entre cultures de conduite des politiques publiques.

« Generally speaking, it can be said that quantitative analyses are especially characteristic of the U.S. as compared to other countries and of younger sociologists as compared to older or impressionist ones. It is a style specifically linked to empiricism and the ideology of progress in the social sciences. And often it is only viewed from this perspective. But as an emphasis and perspective, quantitative style is also associated with political position. The researchers choice of style corresponds to political position regarding the social system and the status quo. » Iverson, 2003

Le point de vue qui nous intéresse ici est celui de l'utilisation. La revendication essentielle des tenants de l'évaluation quantitative est qu'une information plus légitime, objective et prouvée, portée par une communauté scientifique, a vocation à être davantage utilisée et doit conduire à de meilleures utilisations. Dans son approche de l'évaluation centrée sur l'utilisation, Patton (1997) apporte un jugement plus nuancé (encadré 10).

⁷ Qui se traduit, entre bien d'autres exemples, par des titres d'articles scientifiques tels que celui de Holmes et al. (2006), où figurent les mots « déconstruction » et « fascisme ».

Encadré 10

L'évaluation centrée sur l'utilisation

Pour Patton, un critère s'impose pour juger les évaluations – l'utilité : *“utility supplants rigor as the first and most important criterion”*. Les évaluations se distinguent en effet de la recherche par leur objet (des interventions) et par leur public (des décideurs et non des chercheurs). L'accent mis jusque dans les années 1970 sur la rigueur des évaluations aurait rendu les décideurs sceptiques quant à leur utilité réelle car elles se révélaient techniques, coûteuses et parfois difficilement compréhensibles. À partir des années 1980, l'accent est remis sur l'utilité, la rigueur venant au service de celle-ci.

Une « évaluation axée sur l'utilisation » est donc centrée sur son utilisation attendue par les utilisateurs ciblés. Elle est de ce point de vue à la fois personnalisée et contextualisée. L'évaluateur développe une relation de travail avec les utilisateurs ciblés pour déterminer avec eux le type d'évaluation dont ils ont besoin. Ce type d'évaluation ne promeut donc pas une méthode, un modèle, une théorie, un type de données ou un contenu particulier. Elle s'inscrit au contraire dans un processus pour sélectionner ceux-ci de façon adaptée à l'utilisateur et à la situation.

Patton ajoute que les utilisateurs potentiels d'une évaluation seront plus enclins à l'utiliser s'ils comprennent et s'approprient le processus d'évaluation et ses résultats, et donc s'ils y ont été activement associés. En d'autres termes, en associant le futur utilisateur au processus d'évaluation, l'évaluateur prépare les conditions de son utilisation.

Source : Patton (1997).

Toutefois, l'utilité de l'apport de preuves scientifiques n'est pas rejetée par les auteurs constructivistes mettant en avant l'approche argumentative :

« The employment of rational-objectivist techniques may increase the value of argumentation and negotiation – the processes that constitute the roots of policy-oriented learning – through the introduction of systematic evidence. » (Van der Knaap, 1995)

Le cas de l'évaluation des actions de développement peut permettre de discuter cette question. Les évaluations quantitatives ont fait leur apparition dans les pays en développement à partir des années 1990. Ce mouvement a été relayé récem-

ment par une initiative lancée à haut niveau par un *think tank* (le CGDev) et plusieurs grandes fondations. Cette initiative reprend les principes de l'*evidence-based movement* : favoriser l'utilisation de données probantes dans l'élaboration des politiques, promouvoir les évaluations expérimentales et constituer des réservoirs de connaissance scientifique sur des questions pertinentes en matière de politique publique.

Encadré 11

L'initiative 3IE

L'*International Initiative for Impact Evaluation* (3IE) a été lancée en 2007 à la suite du constat d'un déficit de connaissances rigoureuses sur l'impact des programmes de développement et de réduction de la pauvreté (CGDev, 2006). Elle vise à rapprocher :

- les décideurs politiques, et plus particulièrement ceux des pays en développement, à travers le pilotage de l'agenda des évaluations d'impact et l'utilisation de leurs recommandations ;
- les évaluateurs et les chercheurs qui réaliseront des évaluations d'impact, des revues synthétiques et des revues de pairs.

Cette initiative est actuellement soutenue et financée par les gouvernements du Canada, du Danemark, des États-Unis, du Mexique, de l'Ouganda et du Royaume-Uni, ainsi que par CARE, la fondation Bill et Melinda Gates, la fondation William et Flora Hewlett, la fondation des Nations unies, Google.org, l'International Rescue Committee et Save the Children. Elle pourrait contribuer au financement d'une cinquantaine d'évaluations d'impact rigoureuses par an.

Cette initiative s'inscrit pleinement dans la lignée de l'*evidence-based movement* puisqu'elle cherche à promouvoir les politiques et les interventions basées sur les résultats des évaluations d'impact financées.

Une conférence récente organisée par la Banque mondiale a traité la question de l'utilisation des résultats des évaluations quantitatives d'impact dans le domaine du développement⁸. Les conclusions des spécialistes sont mitigées. On peut espérer que la connaissance produite soit plus facilement cumulative, du fait de la proximité des méthodes et des disciplines

⁸ *Making Smart Policy: Using Impact Evaluation for Policy Making*, 15 et 16 janvier 2008.

employées, et que les utilisations conceptuelles soient potentiellement importantes. Mais la diversité des politiques et des contextes, propre au développement, risque de limiter à quelques questions dispersées la création d'un réservoir de connaissances probant.

Pour différentes raisons, l'utilisation instrumentale reste rare : la complexité et la durée de ce type d'évaluation rendent leur appropriation par les décideurs souvent difficile.

« Impact studies generally take considerably longer still, which is one reason why they can seldom be made to dovetail into important policy decisions: rather, they provide long-range feedback on broad strategies (like success of poverty alleviation) more often than on specific policies. » (Cracknell, 2000, p. 205)

Le programme Progesa au Mexique est très souvent cité comme l'illustration par excellence d'une utilisation instrumentale de ce type d'évaluation (encadré 12).

Encadré 12

L'évaluation du programme Progesa

Progesa, aujourd'hui nommé « Oportunidades », est un programme mexicain d'aide sociale qui fournit des allocations aux familles les plus pauvres, à la condition que les enfants aillent à l'école et que les membres de la famille consultent régulièrement au centre de santé.

L'objectif est à la fois d'investir sur le long terme à travers l'éducation et la santé et de réduire sur le court terme l'effet de la pauvreté grâce à l'aide monétaire apportée.

Ce programme, extrêmement novateur, était potentiellement l'objet de nombreuses questions et controverses. La réalisation d'une évaluation rigoureuse et indépendante à un stade précoce de sa mise en œuvre sur des zones pilotes a été identifiée comme un des moyens pour répondre à cette controverse. L'objectif était également de marquer une coupure avec les anciennes pratiques d'arbitrages politiques, en favorisant la transparence et la redevabilité sociale.

L'évaluation d'impact a été mise en œuvre entre 1997 et 2000, selon la méthode d'assignation aléatoire en pipeline des localités desservies. L'impact du programme a été étudié sur 650 variables à partir de plus de 24 000 enquêtes-ménages.

Les résultats positifs de l'évaluation d'impact ont été utilisés par les décideurs politiques. Le changement de gouvernement mexicain entretemps n'a pas empêché le programme Progresa d'être maintenu et étendu, contrairement à la tradition de changement général de politiques à la suite d'élections. L'évaluation a également permis des améliorations opérationnelles dans les règles de Progresa et de proposer une extension aux zones urbaines. À la suite de cette évaluation rigoureuse, ce type de programme a été initié dans 10 nouveaux pays.

Cette évaluation d'impact a également contribué à promouvoir l'évaluation plus systématique des politiques au Mexique et conduit à la création d'un Conseil national d'évaluation de la politique de développement social en 2004 (CONEVAL).

La communication importante faite autour de cet exemple montre que l'utilisation persuasive est potentiellement un effet important de ces évaluations. Dans un domaine comme le développement, dont la légitimité est fragile, il n'est pas étonnant de voir se développer le plaidoyer pour les évaluations quantitatives. Les bailleurs – politiques et contribuables réunis – ont besoin de voir des résultats probants pour poursuivre leur engagement. Dans ce sens, le mouvement actuel correspond à la remise en cause du statu quo « opérationnel » – en revendiquant de se limiter à, et de se concentrer sur ce qui marche de façon prouvée – et à la légitimation de la poursuite du financement du développement lui-même, en établissant de manière objective des résultats probants.

3.7. La faisabilité

Les évaluations sont le plus souvent spécifiquement destinées aux décideurs. La question de la fenêtre d'opportunité politique, déjà évoquée pour la recherche, devient cruciale pour l'utilisation/influence de l'évaluation.

« The alternatives considered in designing, implementing, and assessing a social program are ordinarily those that are within current policy space, the set of alternative policies that can garner political support at any given point of time. »

Un des aspects du travail de l'évaluateur, et sans doute une des valeurs ajoutées des évaluateurs les plus confirmés, sera

Encadré 13

La FAO et les banques de céréales

Les banques de céréales en Afrique ont été initiées par les ONG au milieu des années 1970 au Burkina Faso, puis elles se sont étendues aux autres pays sahéliens dans les années 1980. L'idée était de créer des infrastructures de stockage des céréales dans les villages pour permettre aux paysans de vendre leur récolte au meilleur prix. Elles ont été financées majoritairement à travers des programmes d'aide alimentaire et les ONG. Cependant, il est vite apparu que ces banques de céréales n'empêchaient pas des pertes de stockage importantes et accumulaient des déficits commerciaux, des impayés de prêts et des détournements. Malgré cela, elles ont continué à être massivement soutenues par des subventions jusqu'à la fin des années 1990.

Dès 1991, un rapport d'évaluation commandé par l'USAID (Berg et Kent, 1991) avance des arguments solides pour démontrer leur inefficacité et notamment l'absence de perspective de rentabilité. La réponse des praticiens à ce rapport est défensive, critique ou intentionnellement indifférente. Quatre ans plus tard, une étude de la GTZ confirme le verdict de l'étude de Berg et Kent.

Un atelier est alors organisé en 1996 à Bamako par la GTZ pour rassembler les responsables des différentes organisations engagées dans le financement des banques de céréales, afin d'avoir une discussion ouverte. Les conclusions de cet atelier illustrent le fait qu'un « paradigme de développement » peut avoir la vie dure : elles ne tranchent aucunement entre les tenants d'une poursuite des subventions (ONG notamment) et les opposants au principe du subventionnement continu. En réalité, bien peu des participants à cet atelier avaient connaissance des rapports d'évaluation critiques et la plupart n'étaient pas prêts à écouter les arguments économiques avancés contre le soutien aux banques de céréales puisqu'ils étaient salariés de ces projets. Les tenants d'un abandon des banques de céréales ne sont pas parvenus à faire valoir les conclusions fondées des évaluations, manquant peut-être d'un soutien des décideurs politiques, absents du séminaire.

Bien qu'il y ait eu depuis une réduction des financements aux banques de céréales, ceux-ci ne se tariront sans doute pas tant que l'aide alimentaire en nature aura besoin de déversoirs et que les ONG continueront par idéologie à prôner ce type de schéma.

Source : Reusse (1992).

d'estimer la fenêtre d'opportunité politique dans laquelle ses conclusions peuvent avoir une influence.

Beaucoup d'évaluations jugées sans suite, et bien des frustrations chez les évaluateurs, résultent de conclusions qui sortent tout simplement de l'éventail des positions qu'il semble possible de soutenir de la part des décideurs, compte tenu de l'environnement du moment.

Cette notion de fenêtre politique peut également être la source de dilemmes pour l'évaluateur. Celui-ci peut être partagé entre ce qui lui semble la meilleure recommandation issue de son jugement indépendant (par exemple stopper un programme), tout en sachant que cela n'entre pas dans la fenêtre politique accessible à l'évaluation, et la proposition d'une option sous-optimale (*second-best*) et moins réformatrice (réorienter partiellement le programme) mais plus apte à être utilement transformée en décision.

Ce type de question a donné lieu à un débat au sein de la profession sur la responsabilité de l'évaluateur quant à l'utilisation de son travail (encadré 14).

L'évaluateur a-t-il pour mission de rendre un travail socialement utile, et donc de se plier à la fenêtre politique existante, ou doit-il rendre un jugement souverain et indépendant quels que soient le réalisme et l'opportunité des propositions auxquelles il aboutit ?

« Authors suggest that social acceptability of ideas rather than veracity can result from a strict pragmatic approach, focused on user/buyer needs. » (Garcia, 2008)

Ces questions sont d'autant plus complexes que l'évaluation s'adresse souvent à plusieurs publics possédant chacun des fenêtres politiques différentes : ainsi, les sponsors d'un programme et les acteurs de ces programmes peuvent être tous deux les cibles d'une évaluation, mais leur fenêtre politique respective est fort différente.

L'expérience de l'évaluation de l'aide au développement montre qu'il est en pratique difficile de concilier une préoccupation d'utilisation pour des cibles politiques différentes. Les systèmes d'évaluation qui sont tournés vers les organes de décision politique peinent à être appropriés et utilisés directe-

Encadré 14

L'évaluateur est-il responsable de l'utilisation des évaluations ?

« Is it the responsibility of the evaluator to secure results from evaluation? To get the ball rolling, I will contrast the answers of two mythical evaluation gurus: Michael Patton and Michael Scriven.

First Patton. He might say that evaluation adds value only if its results are used. This means that evaluators must guide decision makers toward the correct decision by helping them identify the right objectives of the program, assess its performance, select the right evidence, identify the proper methods, and so forth... If the evaluator does her job, she will get results. If she does not get results, the chances are she has not done her job as an evaluator. Utilization-based evaluation is quality evaluation. It is the only evaluation worth doing.

By contrast, Scriven might say something like this: to make evaluators responsible for the use of evaluation results is wrong. It leads to supine evaluation, tailored to what clients want to hear. Evaluators have a higher calling: they are mandated to assess the merit and worth of public policies and programs from the standpoint of the citizens, the consumers, the poor, and the voiceless. It is not their responsibility to run programs: to ask them to participate in decision making injures their independence... If it must churn out lessons to justify itself, it will generate an oversupply of simplistic and pious exhortations and platitudes. Worse: evaluators that do not encourage accountability for results fail to provide incentives for learning. »

Source : Piccioto (2004).

ment par les acteurs opérationnels des programmes concernés (Foresti, 2007). De même, les systèmes d'évaluation centrés sur les institutions d'aide atteignent peu les partenaires des pays en développement et sont généralement considérés comme peu utilisés de ce point de vue.

Un aspect important de l'opportunité politique est la dimension temporelle. Les cycles de décisions sont rythmés par des moments particuliers. L'utilisation de l'évaluation, surtout mais pas seulement son évaluation instrumentale, dépend largement de son *timing*.

L'évaluation étant le plus souvent commandée par les décideurs, on pourrait penser que cet ajustement temporel s'effectue assez naturellement. Mais l'anticipation des moments de décision n'est pas si simple. Les cycles du management public semblent souvent se raccourcir.

Une situation d'anticipation insuffisante des moments de décision peut conduire à une incompatibilité avec les cycles normaux de l'évaluation. Le décideur aura alors besoin d'évaluations *in itinere* rapides. Ce dilemme entre le recul (et donc en partie la qualité) pris par l'évaluation et son insertion dans la fenêtre temporelle de la décision se retrouve dans la distinction classique entre évaluation formative et évaluation sommative (encadré 15).

Encadré 15

Évaluations formative et sommative

La distinction entre évaluation formative et évaluation sommative est classique. Ces évaluations se distinguent par le public cible – les acteurs directs dans le cas d'une évaluation formative, un public externe dans le cas d'une évaluation sommative – et par l'objet et le moment de l'évaluation – l'évaluation sommative s'intéresse aux résultats de l'action à évaluer et prend place après que ces résultats ont eu lieu (en principe après la fin de l'action), l'évaluation formative s'intéresse à l'ensemble des aspects d'une action, y compris son processus, et a lieu en cours d'une des phases de l'action lorsque des améliorations sont encore possibles. Ces différences sont bien résumées par la phrase suivante, souvent citée, et attribuée à Robert Stakes : « *When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative* ».

Ces questions temporelles sont particulièrement exigeantes pour les évaluations quantitatives et expérimentales dont la durée normale est de plusieurs années : « *large-scale social experiments that estimate the net effects of major innovative programs may require anywhere from four to eight years to complete and document* » (Rossi et al., 2004).

De fait, des évaluations très rigoureuses peuvent sortir de la fenêtre d'opportunité politique pour des raisons de durée, comme le note Moffitt (2002) dans une revue des évaluations à assignation aléatoire dans le domaine des programmes sociaux aux États-Unis.

« *Those precursors (programs evaluated) were not always, or even mostly, exactly like the post-1996 programs, which weakens the link between the randomized field trials (RFT) findings and current policies. This is a familiar issue with RFTs for experiments take time to conduct and to analyze, and policy interest often has proceeded some distance beyond the programs that were tested by the time the RFT findings are released.* »

Une capacité d'anticipation des besoins de la décision sur plusieurs années n'est pas forcément acquise. Cela peut amener à reconsidérer, pour certains types d'évaluation et dans certains contextes politiques, le bien-fondé de la commande politique de l'évaluation. De même que pour la recherche, on peut plaider pour une identification des thèmes pertinents pour l'évaluation, à partir d'un diagnostic prospectif des besoins mené au sein d'une profession. C'est ce type de démarche que l'initiative 3IE (encadré 10) se propose par exemple de mener.

4. Conclusion

La littérature sur l'utilisation et l'influence de la recherche et de l'évaluation qui a fait l'objet de cette revue est essentiellement théorique. Elle n'est sans doute pas toujours familière aux organisations qui souhaiteraient pouvoir trouver le meilleur équilibre entre connaissances et décisions, et même aux centres de recherche et aux structures d'évaluation qui œuvrent au sein de ces organisations. Pourtant, cette littérature paraît assez directement porteuse d'enseignements pour différents types d'organisations.

Cette conclusion ne propose pas de synthèse des principaux points relevés dans ce qui précède, mais dresse plutôt les principaux enseignements tels que les auteurs de cet article, engagés dans la production et l'utilisation des connaissances au sein d'une organisation d'aide au développement, ont retenu de la littérature passée sous revue dans le cadre de leur expérience.

Le premier de ces enseignements est que l'utilisation de la connaissance dans la décision ne va pas de soi et, par conséquent, qu'elle nécessite une stratégie consciente et affirmée de la part des organisations pour aller au-delà de réussites occasionnelles. Cette stratégie pourrait se résumer sous le slogan : « rendre la connaissance plus influente, rendre la décision plus rationnelle ».

Beaucoup d'éléments pour ce faire ont été évoqués parmi les facteurs déterminants de l'utilisation et de l'influence de la recherche et de l'évaluation. Ils s'adressent à l'ensemble des acteurs concernés :

- les chercheurs et évaluateurs, qui doivent prendre en compte les perspectives d'utilisation tout au long de leur démarche de production de connaissance et passer leur pratique aux tests de la pertinence, de l'opportunité et de la communicabilité ;

- les décideurs, qui seront amenés à anticiper leurs besoins de connaissance, à participer à la commande de recherche et d'évaluation, à partager et débattre les résultats de leurs décisions ;

- les managers, qui peuvent organiser leurs organisations selon les principes des institutions apprenantes, susciter les incitations favorisant une production de connaissance centrée sur l'utilisation, faciliter contacts entre chercheurs, décideurs et opérationnels, et veiller à l'existence d'interfaces efficaces.

Le deuxième enseignement est que les structures dédiées à la production de connaissance appliquée – scientifique ou évaluative – dont l'objet principal est d'être « utile » à la décision, se retrouvent presque inmanquablement dans une situation de porte-à-faux institutionnel, sauf à mettre en place une stratégie explicite (voir *supra*).

Pour la plupart de leurs interlocuteurs – politiques, managers, opérationnels – et même pour l'opinion publique en général, il est avant tout attendu d'elles une utilisation instrumentale de leurs résultats. Or, quelle que soit la configuration de l'organisation, ce type d'utilisation se révèle contingent ou au mieux limité aux décisions courantes portant sur l'amélioration des politiques existantes.

Les producteurs de connaissance peuvent attirer l'attention sur d'autres types d'utilisation ou d'influence de leurs travaux (utilisation conceptuelle, utilité liée au processus, influence indirecte) mais le caractère difficilement démontrable et attribuable de ces mécanismes ne conforte que partiellement leur position. Selon les organisations, cette situation de porte-à-faux peut créer un déficit de légitimité qui se traduit par différents cas de figure :

- la recherche d'ancrage institutionnel. Une procédure rigide de production et de diffusion de la connaissance est une voie pour sécuriser l'utilisation des résultats et asseoir la légitimité des structures en charge. Elle s'écarte cependant de la souplesse nécessaire au schéma des institutions apprenantes. Le risque est la « routinisation » de l'activité au sein de l'organisation, débouchant sur une activité de connaissance se justifiant davantage par son existence même que par son contenu ou son utilité. La faible motivation et le manque d'attractivité qui en découlent sont des caractéristiques que l'on peut parfois retrouver dans les structures d'évaluation ;

- la quête permanente de résultats démontrables. Cette attitude est fréquemment observée dans les structures de production de connaissance à visée décisionnelle. La mise en avant de preuves d'influence (*perceived influence* selon Stone et al., 2001) peut conduire à un certain opportunisme et contribuer parfois à l'auto-attribution de résultats collectifs et diffus, comme l'amélioration des politiques publiques. La recherche de crédibilité peut également favoriser l'utilisation tactique des connaissances en poussant les structures de recherche et d'évaluation à produire des connaissances qui cadrent parfaitement avec l'argumentaire accompagnant les politiques en faveur au moment de la diffusion des résultats ;

- la priorité donnée à des légitimations externes à l'organisation. Les structures de recherche et d'évaluation sont amenées à valider leurs résultats au sein de leur communauté professionnelle. Une priorité trop forte à ce type de validation peut écartier de la préoccupation d'utilisation et d'influence et contribuer à consolider le fossé entre décision et connaissance au sein même de l'organisation.

La question de la légitimité et de la crédibilité des productions de connaissance vis-à-vis des structures qui financent et utili-

sent cette connaissance est un point central des stratégies organisationnelles évoquées ci-dessus.

Le troisième enseignement de cette littérature réside dans l'importance de l'appariement entre processus de production de connaissance et contexte d'utilisation. Cela a particulièrement été mis en valeur dans le cas de l'évaluation dont les différentes modalités ont été discutées – indépendante/ collaborative, quantitative/qualitative, objective/argumentative, somative/formative. La littérature semble associer à chaque contexte – modèle du processus de décision, niveau de décision recherchée, fenêtre politique, *timing* – un instrument d'évaluation privilégié pour favoriser son utilisation. On aurait sans doute pu évoquer le même type de diversité pour les processus de recherche.

Cette diversité soulève un certain nombre de questions. La première est celle de la possibilité de maîtriser et d'utiliser, au sein d'une même organisation, une gamme élargie d'instruments dont certains relèvent de pratiques professionnelles séparées, voire concurrentes. La seconde est la recherche d'un juste équilibre entre l'institutionnalisation de la production de connaissance et son adaptation au contexte. Les dangers d'une institutionnalisation trop forte ont déjà été soulignés. À l'inverse, l'absence de procédures formelles et l'adaptation totale des instruments aux contextes est de nature à marginaliser les processus de production de connaissances et à fragiliser leur place au sein de l'organisation. Ce dilemme entre institutionnalisation et capacité d'adaptation est également une question centrale posée aux stratégies organisationnelles d'utilisation de la connaissance dans la décision.

Bibliographie

- Alkin, M.-C. (1980), "Naturalistic Study of Evaluation Utilization", *New Directions for Program Evaluations: Utilization of Evaluative Information*, vol. 5, Jossey Bass, San Francisco CA.
- Alkin, M.-C. (1990), *Debates on Evaluation*, Sage, Newbury Park, CA.
- Argyris, C. et D. Schön (1978), *Organizational Learning: A Theory of Action Perspective*, McGraw-Hill, New York, NY.
- Banerjee, A., A. Deaton, N. Lustig et K. Rogoff (2006), *An Evaluation of World Bank Research, 1998-2005*, ronéo, Banque mondiale, Washington DC.
- Baumard, P. (1995), « Des organisations apprenantes ? Les dangers de la consensualité », *Revue française de gestion*, Numéro spécial « Les chemins du savoir de l'entreprise », septembre-octobre.
- Baumard, P. (2002), « Géo-économie et stratégie des entreprises », *La Revue française de géo-économie*, vol. 21, n° 2, juin.
- Bellon, B. (2002), « Quelques fondements de l'intelligence économique », *Revue d'économie industrielle*, vol. 98, n° 1.
- Berg, E. et L. Kent (1991), "The Economics of Cereal Banks in the Sahel", *Consultancy report prepared for USAID*, Bethesda, MD.
- Booth, T. (1988), *Developing Policy Research*, Gower Publishing Company Limited, Aldershot.
- Brown, C., L. Daniel et C. Fisher (2006), "From Knowledge Transfer to a Learning-Based Approach: Perspectives from IDS' Information Services", *Knowledge Management for Development Journal*, vol. 2(3).
- Camacho, T.-B. (2007), "Challenges and Key Success Factors to Integrating Learning and Change in Monitoring and Evaluation of Development Projects – Case Study of an Urban Agriculture Project in Eastern Cuba", *Knowledge Management for Development Journal*, vol. 4(1).
- Caplan, N. (1979), "The Two-Communities Theory and Knowledge Utilization", *American Behavioural Scientist*, 23(4).
- Carden, F. (2004), "Les problèmes soulevés par l'évaluation de l'influence politique de la recherche », *Revue internationale des sciences sociales*, n° 179.
- CGDev (2006), *When Will we ever Learn? Improving Lives through Impact Evaluations*, CGDev, Washington, DC.
- Chédotel, F. (2005), « L'improvisation organisationnelle. Concilier formalisation et flexibilité d'un projet », *Revue française de gestion*, vol. 1, n° 154.
- Cohandet, P. et P. Llerena (1999), « La conception de la firme comme processeur de connaissances », *Revue d'économie industrielle*, vol. 88, n° 1.
- Cook, T.-D. (1997), "Lessons Learned in Evaluation over the past 25 years", in Chelimsky, E. et W.-R. Shadish (dir. pub.), *Evaluation for the 21st Century: A Handbook*, SAGE Publications, Inc., Thousand Oaks, CA.

- Cors, R. (2003), *What is a Learning Organization? Reflections on the Literature and Practitioner Perspectives*, University of Wisconsin, Madison, WI.
- Cousins, J.-B. (2004), "Crossing the Bridge: Toward Understanding Use through Systematic Inquiry", in Alkin, M.-C. (dir. pub.), *Evaluation Roots*, SAGE Publications, Inc., Thousand Oaks, CA.
- Cousins, J.-B. (2006), "Non Academic Impact of Research through the Lens of Recent Developments in Evaluation", intervention lors de la conférence "New Frontiers in Evaluation", avril, Vienne.
- Cracknell, B.-E. (2000), *Evaluating Development Aid: Issues, Problems and Solutions*, SAGE Publications, Inc., Thousand Oaks, CA.
- Cyert, R. et J. March (1963), "A Behavioural Theory of the Firm", *Prentice Hall, Upper Saddle River, NJ*.
- Dolowitz, D. et D. Marsh (1996) "Who Learns What from Whom: a Review of the Policy Transfer Literature", *Political Studies*, XLIV.
- Fairhead, J. et M. Leach (1996), "Enriching the Landscape: Social History and the Management of Transition Ecology in the Forest. Savanna Mosaic of the Republic of Guinea", *Journal of the International African Institute*, vol. 66, n° 1.
- Ferguson, J.-E. (2005), "Bridging the Gap between Research and Practice", *Knowledge Management for Development Journal*, vol. 1(3).
- Foray, D. (1993), « Recherche et technologie – Autour de l'apprentissage organisationnel et de l'économie du savoir », *Revue d'économie industrielle*, vol. 65, n° 1.
- Foray, D. (1998), « Science, technologie et marché », *Working Papers IMRI*, n° 98/09, université Paris-Dauphine, Paris.
- Foresti, M. (2007), *A Comparative Study of Evaluation Policies and Practices in Development Agencies*, série Notes méthodologiques, Ex-post n° 1, AFD, Paris.
- Furubo, J.-E. (1994), "Learning from Evaluations: The Swedish Experience", in Leeuw, F., R. Rist et R. Sonnichsen (dir. pub.), *Can Government Learn? Comparative Perspectives on Evaluation & Organizational Learning*, Transaction Publishers, New Brunswick, NJ.
- Furubo, J.-E. (2005), in Rist and Stame (2005).
- GAO (1992), *The Evaluation Synthesis*, GAO/PEMD, 10.1.2, United States General Accounting Office, Washington, D.C. (disponible sur : <http://www.gao.gov/special.pubs/pemd1012.pdf>).
- Garcia, J. (2008), "A Realistic Account of Evidence-Informed Tobacco Control Practice in Ontario Public Health Agencies", Thèse de doctorat en philosophie, University of Waterloo, Ontario.
- Grindle, M.-S. (dir. pub.) (1997), *Getting Good Government: Capacity Building in the Public Sector of Developing Countries*, Harvard Institute for International Development, Cambridge, MA.
- Grindle, M.-S. et J.-W. Thomas (1990), "After the Decision: Implementing Policy Reform in Developing Countries", *World Development*, 18 (8).
- Guba, E.-G. et Y.-S. Lincoln (1989), *Fourth Generation Evaluation*, SAGE Publications, Inc., Thousand Oaks, CA.
- Hanney, S.-R. (2004), "Personal Interaction with Researchers or Detached Synthesis of Evidence: Modelling the Health Policy Paradox", *Evaluation and Research in Education*, vol. 18, n° 1&2.

- Hanney, S.-R., M.-A. Gonzales, M.-J. Buxton et M. Kogan (2003), "The Utilization of Health Research in Policy-Making: Concepts, Examples and Methods of Assessment", *Health Policy Research and Systems*, 1(2).
- Hassenteufel, P. et A. Smith (2002), « Essoufflement ou second souffle ? L'analyse des politiques publiques "à la française" », *Revue française de science politique*, vol. 52, n° 1.
- Hatchuel, A. (1999), « Connaissances, modèles d'interaction et rationalisations – De la théorie de l'entreprise à l'économie de la connaissance », *Revue d'économie industrielle*, vol. 88(1).
- Hennink, M. et R. Stephenson (2004), "Using Research to Inform Health Policy : Barriers and Strategies in Developing Countries", *Opportunities and Choices Working Paper*, n° 9, University of Southampton, Southampton.
- Henry, G.-T. (2003), "Influential Evaluations", *American Journal of Evaluation*, vol. 24, n° 4.
- Henry, G.-T. et M.-M. Mark (2003), "Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions", *American Journal of Evaluation*, vol. 24, n° 3.
- Henry, G.-T. et M.-M. Mark (2004), "The Mechanisms and Outcomes of Evaluation Influence", *American Journal of Evaluation*, vol. 10, n° 1.
- Holmes, D., S.-J. Murray, A. Perron et G. Rail (2006), "Deconstructing the Evidence-based Discourse in Health Sciences: Truth, Power and Fascism", *Journal of Health Politics, Policy and Law*, 2006:4.
- Huber, G.-P. (1991), "Organizational Learning: The Contributing Processes and the Literatures", *Organization Science*, vol. 2, n° 1.
- Innvaer, S., G. Vist, M. Trommald et A. Oxman (2002), "Health Policy Makers' Perceptions of Their Use of Evidence: A Systematic Review", *Journal of Health Services Research and Policy* 7.
- Iverson, A. (2003), *Attribution and Aid Evaluation in International Development: A Literature Review*, ronéo, Evaluation Unit, IDRC, Ottawa.
- Jacob, J.-P. (2000), « Connaissance et développement en Afrique », *Sciences sociales et coopération en Afrique. Les rendez-vous manqués*, in Jacob, J.-P. (dir. pub.), *Nouveaux Cahiers de l'IUED*, n° 10, Presses universitaires de France/IUED, Paris/Genève.
- Kennedy, M.-M. (1983), "Working Knowledge", *Knowledge: Creation, Diffusion, Utilization*, 5.
- Kirkhart, K.-E. (2000) "Reconceptualizing Evaluation Use: An Integrated Theory of Influence", in Caracelli, V.-J. et H. Preskill (dir. pub.), "The Expanding Scope of Evaluation Use", *New Directions for Evaluation* 88, Jossey-Bass, San Francisco, CA.
- Krugman, P. (1997), *Development, Geography, and Economic Theory*, MIT Press, Cambridge, MA.
- Lavis J.-N., D. Robertson, J.-M. Woodside, C.-B. MacLeod et J. Abelson (2003), "How Can Research Organizations more Effectively Transfer Research Knowledge to Decision Makers ?", *The Milbank Quarterly*, 81(2).
- Machlup, F. (1993), « Économie des connaissances et de l'information », *Réseaux. Communication – Technologie – Société*, vol. 11(58).
- MacLure, M. (2005). "Clarity Bordering on Stupidity: Where's Quality in Systematic Review?", *Journal of Education Policy*, vol. 20, n° 4.

- Moffitt, R. (2002), "The Role of Randomized Field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs", *NBER Working Paper* n° T0295
- Muller, P. (2005), « Esquisse d'une théorie du changement dans l'action publique. Structures, acteurs et cadres cognitifs », *Revue française de science politique*, vol. 55, n° 1.
- Neilson, S. (2001), *Knowledge Utilization and Public Policy Processes: A Literature Review*, IDRC-Supported Research and its Influence on Public Policy, Evaluation Unit, IDRC, Ottawa.
- Nonaka, I. et H. Takeuchi (1995), *The Knowledge-Creating Company*, Oxford University Press, New York, NY.
- Nutley, S., I. Walter et H. Davies (2003), "From Knowing to Doing: A Framework for Understanding the Evidence-Into-Practice Agenda", *Evaluation*, vol. 9(2), Sage Publications.
- Patton, M.-Q. (1978), *Utilization-focussed Evaluation*, Sage, Beverly Hills, CA.
- Patton, M.-Q. (1997), *Utilization-focussed Evaluation*. The New Century Text, 3e édition, SAGE Publications, Inc., Thousand Oaks, CA.
- Pawson, R. (2006), *Evidence-based Policy: a Realist Perspective*, Sage, Londres.
- Pénin, J. (2003), « Endogénéisation des externalités de la recherche : le rôle de la capacité d'émission des connaissances », *Revue d'économie industrielle*, vol. 102(1).
- Philpott, A. (1999), *Twists in the Mwanza Tale: Did one HIV Research Study Shift Global Policy?* Institute of Development, Londres (disponible sur <http://nt1.ids.ac.uk>).
- Piccio, R. (2004), "Use of Evaluation Findings to Improve Development Effectiveness: Panel Discussion?", in Pitman, G.-K., O.-N. Feinstein et G.-K. Ingram (dir. pub.), *Evaluating Development Effectiveness*, World Bank Series on Evaluation and Development, Transaction Publishers, New Brunswick et Londres.
- Porter, R. et I. Hicks (1995), *Knowledge Utilization and the Process of Policy Formulation: Towards a Framework for Africa*, USAID, Washington, DC.
- Reusse, E. (1992), *The Ills of Aid. An Analysis of Third World Development Policies*, The University of Chicago Press, Chicago.
- Rey, O. (2006), « Qu'est-ce qu'une bonne' recherche en éducation ? », *Lettre d'information de la cellule veille scientifique et technologique*, n° 18.
- Rist, R. et N. Stame (dir. pub.) (2005), "From Studies to Streams: Managing Evaluative Systems", *Comparative Policy Evaluation*, vol. 12, Transaction Publishers, Londres.
- Rossi, P.-H., H.-E. Freeman et M.-W. Lipsey (2004), *Evaluation: A Systematic Approach* (septième édition), SAGE Publications, Inc., Thousand Oaks, CA.
- Scheerens, J. (2001), "The Concept of Learning Organization and the Evaluation Function", document préparé pour le projet DESLO, université de Twente, Enschede.
- Shulha, L.-M. et J.-B. Cousins (1997), "Evaluation Use: Theory, Research and Practice since 1986", *Evaluation Practice* 18.

- Stone, D., S. Maxwell et M. Keating (2001), *Bridging Research and Policy*, intervention lors d'un atelier international financé par le DFID, Warwick University, 16-17 juillet (disponible sur www.gdnet.org/pdf/Bridging.pdf).
- Sutton, R. (1999), "The Policy Process: An Overview", *Working Paper 118*, ODI, Londres.
- Valovirta, V. (2002), "Evaluation Utilization as Argumentation", *Evaluation*, 8(1).
- Van der Knaap, P. (1995), "Policy Evaluation and Learning – Feedback, Enlightenment or Argumentation?", *Evaluation*, vol. 1(2), Sage Publications, Londres.
- Van Winkelen, C. et J. McKenzie (2006), "Integrating Individual and Organizational Learning Initiatives: Working across Knowledge Management and Human Resource Management Functional Boundaries", *The Electronic Journal of Knowledge Management*, vol. 5(4).
- Weiss, C. (1979), "The Many Meanings of Research Utilization", *Public Administration Review*, vol. 39, n° 5.
- Weiss, C. (1986), "The Circuitry of Enlightenment", *Knowledge: Creation, Diffusion, Utilization*, vol. 8, n° 2, SAGE Publications, Inc., Thousand Oaks, CA.
- Weiss, C. (1998), "Have we Learned Anything New about the Use of Evaluation?", *American Journal of Evaluation* 19(1).
- Wiedenhof, C. et H. Molenaar (2006), "One never Knows. Research Policy and Knowledge Management in Dutch Development Cooperation", *Knowledge Management for Development Journal*, vol. 2(3).